

Лабораторная работа №1

Алгоритмы сжатия: Шеннона-Фано, Хаффмана, арифметический, словарно-ориентированные алгоритмы

Цель работы: изучить алгоритмы сжатия информации.

Задачи работы:

- изучить основные определения и теоремы сжатия информации;
- изучить приведенные в лабораторной работе примеры;
- выполнить индивидуальное задание и защитить у преподавателя.

1. Краткая теория

Характерной особенностью большинства типов данных является их избыточность.

Степень избыточности данных зависит от типа данных.

Например, для видеоданных степень избыточности в несколько раз больше чем для графических данных, а степень избыточности графических данных, в свою очередь, больше чем степень избыточности текстовых данных.

Другим фактором, влияющим на степень избыточности, является принятая система кодирования.

Примером систем кодирования могут быть обычные языки общения, которые являются ни чем другим, как системами кодирования понятий и идей для высказывания мыслей. Так, установлено, что кодирование текстовых данных с помощью средств русского языка дает в среднем избыточность на 20-25% большую, чем кодирование аналогичных данных средствами английского языка.

Для человека избыточность данных часто связана с качеством информации, поскольку избыточность, как правило, улучшает понятность и восприятие информации. Однако, когда речь идет о хранении и передаче информации средствами компьютерной техники, то избыточность играет отрицательную роль, поскольку она приводит к возрастанию стоимости хранения и передачи информации. Особенно актуальной эта проблема становится в случае обработки огромных объемов информации при незначительных объемах носителей данных. В связи с этим, постоянно возникает проблема уменьшения избыточности или сжатия данных.

Все алгоритмы сжатия оперируют входным потоком информации, минимальной единицей которой является бит, а максимальной – несколько бит, байт или несколько байт. Целью процесса сжатия есть получение более компактного выходного потока информационных единиц из изначально некомпактного входного потока при помощи их преобразования. Основными техническими характеристиками процессов сжатия и результатов их работы являются:

- степень сжатия (compress rating) или отношение (ratio) объемов исходного и результирующего потоков;

– скорость сжатия – время, затрачиваемое на сжатие некоторого объема информации входного потока, до получения из него эквивалентного выходного потока;

– качество сжатия – величина, показывающая, на сколько сильно упакован выходной поток, при помощи применения к нему повторного сжатия по этому же или иному алгоритму.

Все способы сжатия можно разделить на две категории: обратимое и необратимое сжатие.

Под необратимым сжатием подразумевают такое преобразование входного потока данных, при котором выходной поток представляет достаточно похожий по внешним характеристикам на входной поток объект, однако отличается от него объемом. Степень сходства входного и выходного потоков определяется степенью соответствия некоторых свойств объекта, представляемого данным потоком информации.

Обратимое сжатие всегда приводит к снижению объема выходного потока информации без изменения его информативности, т.е. без потери информационной структуры. Из выходного потока, при помощи восстанавливающего или декомпрессирующего алгоритма, можно получить входной, а процесс восстановления называется декомпрессией или распаковкой и только после процесса распаковки данные пригодны для обработки в соответствии с их внутренним форматом.

В обратимых алгоритмах кодирование, как процесс, можно рассматривать со статистической точки зрения, что еще более полезно не только для построения алгоритмов сжатия, но и для оценки их эффективности.

Для всех обратимых алгоритмов существует понятие стоимости кодирования. Под стоимостью кодирования понимается средняя длина кодового слова в битах. Избыточность кодирования равна разности между стоимостью и энтропией кодирования, а хороший алгоритм сжатия всегда должен минимизировать избыточность. Для любого алгоритма, всегда имеется некоторый предел степени сжатия, определяемый энтропией входного потока.

Алфавитное кодирование

Алфавитное кодирование задается схемой, в которой каждой букве алфавита ставится в соответствие двоичная последовательность символов:

$$f : \begin{cases} b_1 \rightarrow v_1 \\ b_2 \rightarrow v_2 \\ \dots \\ b_m \rightarrow v_m, \end{cases} \quad v_i \in \{0,1\}^*.$$

Коды v_i называются элементарными, а их набор $V=(v_1, v_2, \dots, v_m)$ – кодом.

Пример №1

Рассмотрим схему

$$f_1: \begin{cases} b_1 \rightarrow 0 \\ b_2 \rightarrow 01. \end{cases}$$

f_1 задает взаимно-однозначное кодирование. Перед каждым вхождением символа 1 стоит символ 0, поэтому каждое вхождение 1 вместе с предшествующим нулем кодирует букву b_2 . Символ 0, за которым не следует 1, кодирует букву b_1 .

Последовательность $b = 001010001$ имеет единственную расшифровку $a = b_1 b_2 b_2 b_1 b_1 b_2$.

Пример №2

Рассмотрим схему

$$f_2: \begin{cases} b_1 \rightarrow 0 \\ b_2 \rightarrow 01 \\ b_3 \rightarrow 001. \end{cases}$$

Кодирование, задаваемое схемой f_2 , не является взаимно-однозначным.

Последовательность $b=001$ допускает две расшифровки $a_1=b_1 b_2$ и $a_2=b_3$.

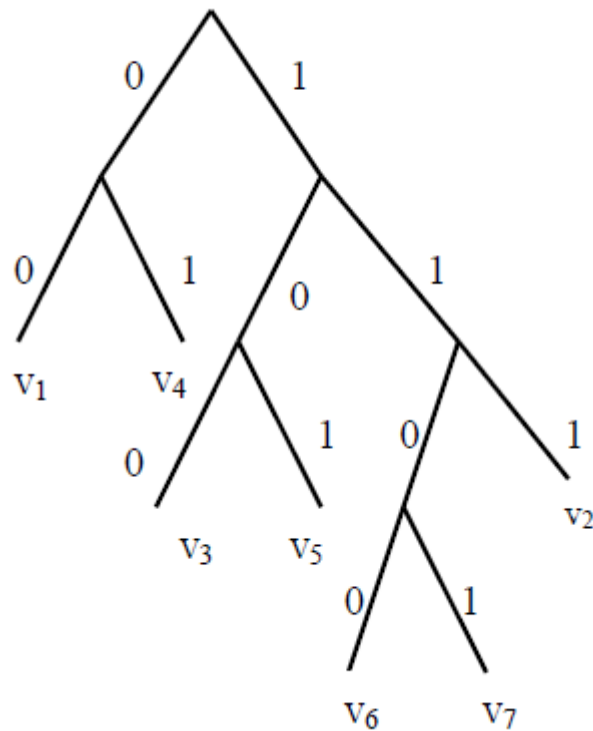
Для непосредственной проверки взаимной однозначности необходимо в общем случае проверить бесконечное множество пар слов.

Пусть задан код $V=(v_1, v_2, \dots, v_m)$. Элементарные коды определяют бинарное кодовое дерево. Из каждой вершины дерева выходит не более двух ребер в вершины следующего яруса, левое из которых помечается символом 0, а правое – символом 1. Элементарным кодам соответствуют вершины дерева, определяемые путем, идущим от корня. Если код префиксный, элементарные коды расположены в листьях дерева.

Дерево называется насыщенным, если из каждой вершины, не являющейся листом, в следующий ярус выходит ровно два ребра.

Дерево для схемы префиксного кодирования f_3 .

$$f_3: \begin{cases} b_1 \rightarrow 00 \\ b_2 \rightarrow 111 \\ b_3 \rightarrow 100 \\ b_4 \rightarrow 01 \\ b_5 \rightarrow 101 \\ b_6 \rightarrow 1100 \\ b_7 \rightarrow 1101. \end{cases}$$



Алгоритм Хаффмана

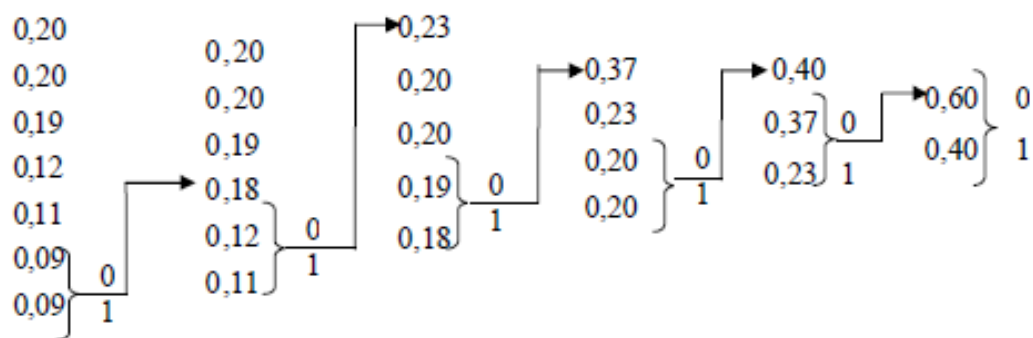
Алгоритм Хаффмана строит оптимальный префиксный код. При рассмотрении алгоритмов кодирования наряду с элементарными кодами вершинам кодового дерева будем приписывать вероятности соответствующих букв. Алгоритм Хаффмана основан на следующих свойствах оптимальных кодов.

Теорема редукции позволяет свести задачу построения оптимального кода мощности m к задаче построения оптимального кода мощности $m-1$. На ней основан алгоритм Хаффмана, который заключается в следующем. Пусть вероятности в распределении $P=(p_1, p_2, \dots, p_m)$ расположены в порядке невозрастания. На каждом шаге объединяются две буквы, имеющие наименьшие вероятности. Вместо этих двух букв вводится новая буква с вероятностью $p=p_{m-1}+p_m$. Вероятность p вставляется в оставшийся набор вероятностей так, чтобы в получившемся новом наборе вероятности остались расположенными в порядке невозрастания. Продолжаем процесс объединения вероятностей до тех пор, пока не останутся две буквы алфавита. Одной из них приписывается символ 0, другой – символ 1 (оптимальный код для двух букв при произвольном распределении вероятностей). Затем из оптимального кода для двух букв строится оптимальный код для трех букв, и т.д. Продолжая этот процесс, придем к искомому оптимальному коду для m букв.

Пример №3.

Пусть $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,20;0,20;0,19;0,12;0,11;0,09;0,09)$.

Процесс построения оптимального кода можно представить следующим образом:



Фигурными скобками отмечены объединяемые вероятности. Для каждой скобки верхнему члену приписываем символ 0, нижнему – символ 1. Затем осуществляем движение в обратном направлении к p_1, p_2, \dots, p_7 и, проходя скобки, выписываем соответствующие элементарные коды.

Например, путь 0,60 – 0,23 – 0,11 дает элементарный код 011 для буквы b_5 . Таким образом, мы получаем следующую схему f для оптимального кода:

$$f : \begin{cases} b_1 \rightarrow 10 \\ b_2 \rightarrow 11 \\ b_3 \rightarrow 000 \\ b_4 \rightarrow 010 \\ b_5 \rightarrow 011 \\ b_6 \rightarrow 0010 \\ b_7 \rightarrow 0011 \end{cases}$$

Стоимость кодирования $C^*(P)=2,78$.

Алгоритм Фано

Упорядоченный в порядке невозрастания вероятностей список букв делится на две последовательные части так, чтобы суммы вероятностей входящих в них букв как можно меньше отличались друг от друга. Буквам из первой части приписываем символ 0, а буквам из второй части – символ 1. Далее точно так же поступаем с каждой из полученных частей, если она содержит хотя бы две буквы. Построенный код является префиксным, и ему соответствует насыщенное кодовое дерево.

В алгоритме Фано кодовое дерево строится от корня, а в алгоритме Хаффмана – начиная с листьев. Это отличие позволяет в алгоритме Хаффмана полнее использовать специфику данного распределения вероятностей и строить оптимальный код. Алгоритм Фано строит код, близкий к оптимальному.

Пример №4. Применим алгоритм Фано к тому же распределению вероятностей.

Пусть $V = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P = (0,20; 0,20; 0,19; 0,12; 0,11; 0,09; 0,09)$.

0,59	{ 0,20 0,20 0,19 }	0	0,39	{ 0,20 0,20 0,19 }	0	0,20	0
						0,20	0
						0,19	1
0,41	{ 0,12 0,11 0,09 0,09 }	1	0,23	{ 0,12 0,11 }	0	0,12	0
						0,11	1
			0,18	{ 0,09 0,09 }	1	0,09	0
						0,09	1

Получаем следующую схему алфавитного кодирования:

$$f : \begin{cases} b_1 \rightarrow 00 \\ b_2 \rightarrow 010 \\ b_3 \rightarrow 011 \\ b_4 \rightarrow 100 \\ b_5 \rightarrow 101 \\ b_6 \rightarrow 110 \\ b_7 \rightarrow 111. \end{cases}$$

Стоимость кодирования $C_F(P) = 2,80$.

Алгоритм Шеннона

Алгоритм Шеннона применим в случае, когда все вероятности $p_i > 0$. Букве b_i ставится в соответствие последовательность из $l_i = \left\lceil \log \frac{1}{p_i} \right\rceil$ двоичных символов (здесь $\lceil x \rceil$ – ближайшее целое сверху числа x и \log здесь и везде далее берется по основанию 2). Алгоритм Шеннона основан на том, что выбранные длины l_i ($i=1, 2, \dots, m$) удовлетворяют неравенству Мак-Миллана. После выбора длин применяется алгоритм Шеннона построения схемы кодирования по заданному набору длин элементарных кодов, описанный ранее.

Пример №5.

Пусть $V = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P = (0,20; 0,20; 0,19; 0,12; 0,11; 0,09; 0,09)$.

Вычислим набор длин для P .

$$l_1 = l_2 = \left\lceil \frac{1}{0,20} \right\rceil = 3,$$

$$l_3 = \left\lceil \frac{1}{0,19} \right\rceil = 3,$$

$$l_4 = \left\lceil \frac{1}{0,12} \right\rceil = 4,$$

$$l_5 = \left\lceil \frac{1}{0,11} \right\rceil = 4,$$

$$l_6 = l_7 = \left\lceil \frac{1}{0,09} \right\rceil = 4.$$

Построим префиксный код по алгоритму Шеннона с вычисленными длинами элементарных кодов.

$$f: \begin{cases} b_1 \rightarrow 000 \\ b_2 \rightarrow 001 \\ b_3 \rightarrow 011 \\ b_4 \rightarrow 1001 \\ b_5 \rightarrow 1011 \\ b_6 \rightarrow 1101 \\ b_7 \rightarrow 1110. \end{cases}$$

Стоимость кодирования $C_{\text{Ш}}(P) = 3,41$.

Рассмотрим строку «beer boor beer!». Чтобы сконструировать кодовые значения для всех элементов в зависимости от их частоты, нужно построить такое бинарное дерево, в котором каждый лист содержит символ из данной строки. Дерево мы необходимо строить от листьев к корню, чтобы элементы с меньшей частотой находились дальше от корня, а более часто встречающиеся — ближе к нему.

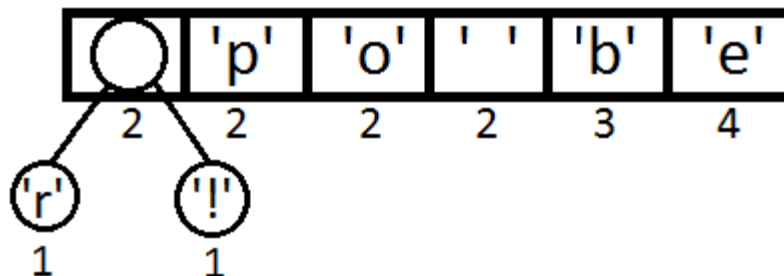
Для строительства такого дерева воспользуемся очередью с приоритетом, но немного модифицированной — инвертируем приоритеты, чтобы элемент с наименьшим приоритетом был самым важным. Получается, очередь сначала будет возвращать элементы с наименьшей частотой. Эта модификация поможет нам построить дерево начиная с листьев.

Для начала посчитаем частоту каждого символа:

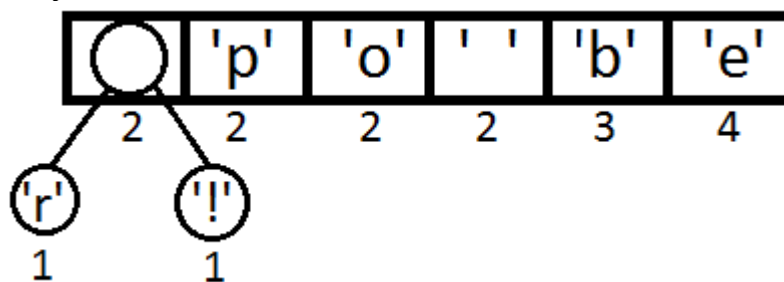
Символ	Частота
‘b’	3
‘e’	4
‘p’	2
‘ ‘	2

'o'	2
'r'	1
'!'	1

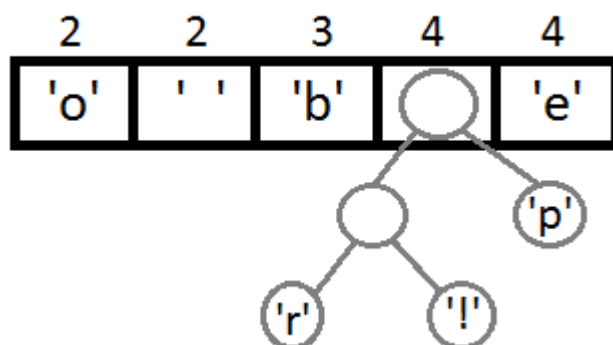
После этого создаются элементы бинарного дерева для каждого символа и представляются как очередь с приоритетом, в качестве которого используется частота.

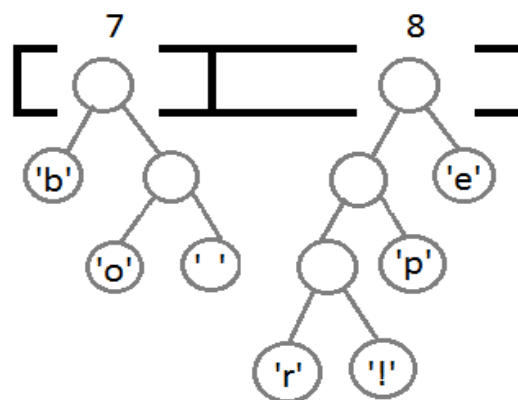
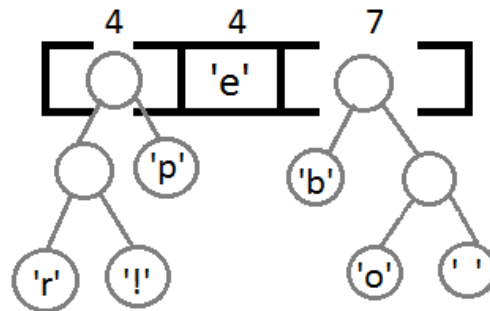
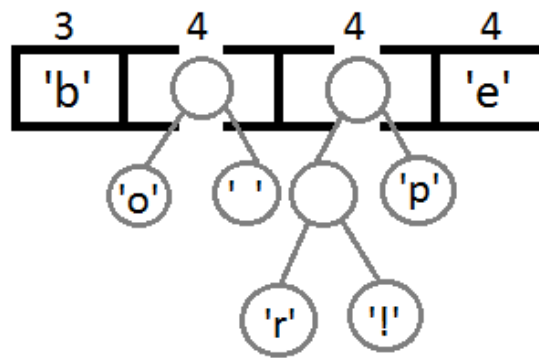


Теперь возьмем первые два элемента из очереди и создадим третий, который будет их родителем. Этот новый элемент поместим в очередь с приоритетом, равным сумме приоритетов двух его потомков. Иначе говоря, равным сумме их частот.

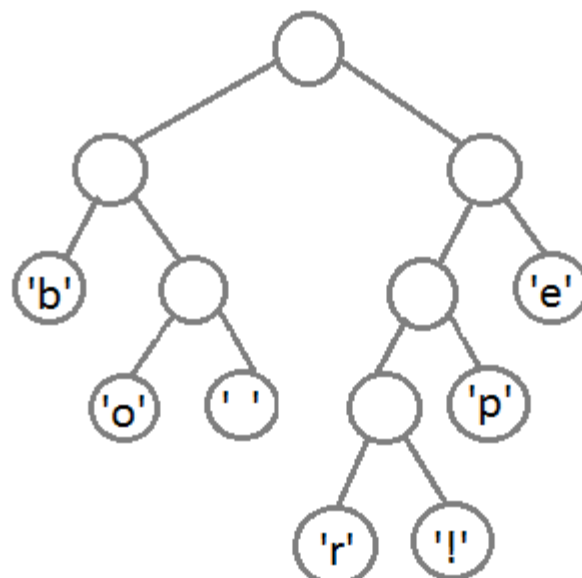


Далее будем повторять шаги, аналогичные предыдущему:

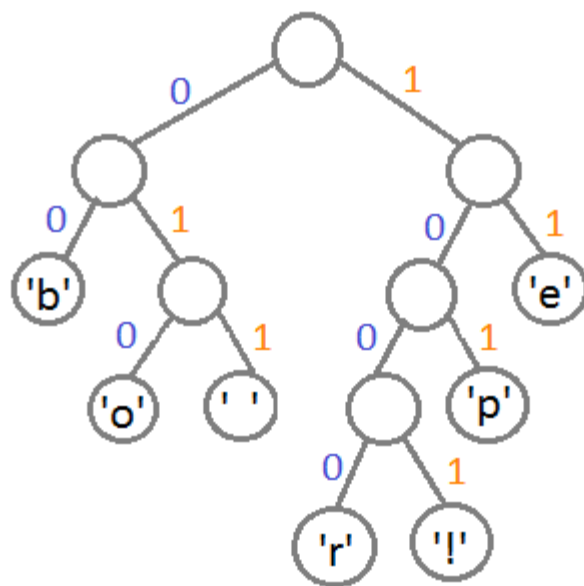




Теперь, после объединения последних двух элементов с помощью их нового родителя, мы получим итоговое бинарное дерево:



Осталось присвоить каждому символу его код. Для этого запустим обход в глубину и каждый раз, рассматривая правое поддерево, будем записывать в код 1, а рассматривая левое поддерево — 0.



В результате соответствие символов кодовым значениям получится следующим:

Символ	Кодовое значение
'b'	00
'e'	11
'p'	101
' '	011
'o'	010
'r'	1000
'!'	1001

Декодировать последовательность битов очень просто: нужно обходить дерево, отбрасывая левое поддерево, если встретилась единица и правое, если встретился 0. Продолжать обход нужно до тех пор, пока не встретим лист, т.е. искомое значение закодированного символа. Например, закодированной строке «101 11 101 11» и нашему дереву декодирования соответствует строка «рере».

Здесь важно заметить, что ни одно кодовое значение не является префиксом никакого другого. В нашем примере 00 — код символа «b». Это значит, что 000 уже не может быть кодом ни для какого другого символа, т.к. иначе это привело бы к конфликту при декодировании. Ведь если после прохода по 00 мы пришли к листу «b», то по кодовому значению 000 мы уже не найдем никакого другого символа.

Задания для самостоятельной работы:

Вариант №1

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,3;0,1;0,1; 0,15; 0,2; 0,15\}$.
2. Закодировать предложение: Я учусь на факультете прикладной информатики.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,20;0,19;0,14;0,19;0,28)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №2

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,1;0,22;0,4; 0,2; 0,08\}$.
2. Закодировать предложение: Моя будущая специальность – информационные системы и технологии.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,001;0,5;0,18;0,299;0,02)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №3

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4\}$, $P=\{0,12;0,12;0,06; 0,7\}$.
2. Закодировать предложение: ФГБОУ ВПО «Кубанский государственный аграрный университет».
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,25;0,119;0,14;0,03;0,01; 0,2; 0,251)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №4

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,02; 0,02; 0,1; 0,07; 0,3; 0,06; 0,13; 0,4\}$.
2. Закодировать предложение: Куратор группы – Орлянская Наталья Петровна.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,25;0,119;0,14;0,03;0,01; 0,2; 0,261)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №5

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=\{0,02; 0,03; 0,03; 0,03; 0,7; 0,09; 0,1\}$.
2. Закодировать предложение: За факультетом закреплено общежитие №20.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,1;0,15;0,4;0,0009;0,1; 0,2; 0,0491)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №6

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,07; 0,1; 0,08; 0,08; 0,67\}$.
2. Закодировать предложение: Город Краснодар столица Краснодарского края.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,11;0,3;0,06; 0,4; 0,13)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №7

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,021; 0,1; 0,08; 0,08; 0,4; 0,06; 0,17; 0,089\}$.
2. Закодировать предложение: Краснодарский край входит в состав Южного Федерального округа.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,1;0,05;0,22;0,07;0,1; 0,12; 0,34)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №8

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,02; 0,04; 0,18; 0,7; 0,03; 0,03\}$.
2. Закодировать предложение: Кубанский госагроуниверситет – вуз с богатейшей историей, прогрессивным будущим и надежным будущим.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,05;0,11;0,4;0,2;0,07; 0,002; 0,168)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №9

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,01; 0,01; 0,13; 0,25; 0,25; 0,06; 0,18; 0,11\}$.

2. Закодировать предложение: Факультет прикладной информатики организован в 2000 году для подготовки специалистов в области информационных технологий.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,13;0,08;0,24;0,05;0,1; 0,1; 0,3)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №10

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,25; 0,01; 0,18; 0,07; 0,15; 0,04; 0,1; 0,2\}$.

2. Закодировать предложение: По окончании бакалавриата желающие имеют возможность продолжить обучение в магистратуре, а затем и в аспирантуре.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,22; 0,7;0,4;0,5;0,11; 0,3; 0,3)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №11

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,4;0,12;0,12; 0,13; 0,09; 0,14\}$.

2. Закодировать предложение: Заместитель декана по воспитательной работе: старший преподаватель Гусельникова Александра Алексеевна.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,13;0,05;0,06;0,5;0,26)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №12

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,51;0,03;0,1; 0,06; 0,3\}$.

2. Закодировать предложение: Заместитель декана по учебной работе: доцент Василенко Игорь Иванович.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,1;0,25;0,05;0,05;0,55)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №13

Проведите кодирование следующих последовательностей символов.

1. Построить префиксный код: $N=(1,1,4,3,7,7,3,4)$.

2. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4\}$, $P=\{0,4;0,1;0,3; 0,2\}$.

3. Закодировать предложение: Декан: профессор Курносков Сергей Андреевич.

4. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,3;0,069;0,27;0,02;0,02; 0,1; 0,221)$

5. Для заданий 1-3 определить стоимость кодирования.

Вариант №14

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,1; 0,12; 0,11; 0,17; 0,03; 0,6; 0,18; 0,15\}$.

2. Закодировать предложение: Студенты получают образование в области информатики, подготовку для работ по созданию, внедрению и сопровождению информационных систем.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,3;0,15;0,02;0,17;0,1; 0,12; 0,14)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №15

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=\{0,26; 0,01; 0,13; 0,2; 0,1; 0,2; 0,1\}$.

2. Закодировать предложение: «Прикладная информатика», «Бизнес-информатика» и «Информационные системы и технологии».

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,11;0,4;0,3;0,08;0,1; 0,005; 0,005)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №16

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,16; 0,08; 0,3; 0,2; 0,404\}$.

2. Закодировать предложение: Новое направление подготовки ИТ-специалистов в области информационных систем автоматизации бизнес-процессов и управления ими.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,05;0,2;0,01; 0,5; 0,24)$

4. Для заданий 1-3 определить стоимость кодирования.

Вариант №17

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,12; 0,15; 0,07; 0,07; 0,16; 0,03; 0,3; 0,1\}$.

2. Закодировать предложение: Специальность уникальна тем, что студенты одновременно получают два образования - в области информатики и в области экономики.

3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,1;0,05;0,22;0,07;0,1; 0,52; 0,04)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №18

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,02; 0,13; 0,03; 0,3; 0,5; 0,02\}$.
2. Закодировать предложение: Специальность направлена на подготовку профессионалов в области разработки, создания и эксплуатации комплексных интегрированных информационно-управляющих систем.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,13;0,01;0,25;0,3;0,01; 0,1; 0,2)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №19

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,19; 0,21; 0,03; 0,16; 0,16; 0,05; 0,1; 0,1\}$.
2. Закодировать предложение: Центр предвузовской подготовки находится в корпусе факультета защиты растений, аудитория № 316.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,1;0,1;0,32;0,07;0,16; 0,16; 0,09)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №20

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,15; 0,01; 0,12; 0,3; 0,02; 0,01; 0,19; 0,2\}$.
2. Закодировать предложение: Управление науки и инноваций (УНИ) является структурным подразделением Кубанского государственного аграрного университета.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,24; 0,17;0,1;0,03;0,15; 0,01; 0,3)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №21

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,7;0,03;0,12; 0,01; 0,01; 0,13\}$.
2. Закодировать предложение: Студенческое проектно-конструкторское бюро (СПКБ) начало свою деятельность в 1970 году.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,21;0,14;0,14;0,21;0,3)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №22

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,4;0,17;0,03; 0,3; 0,1\}$.
2. Закодировать предложение: Малые инновационные предприятия (далее МИП) являются важным составляющим компонентом инновационной системы.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,23;0,5;0,07;0,1;0,1)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №23

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4\}$, $P=\{0,2;0,2;0,1; 0,5\}$.
2. Закодировать предложение: МИПы при вузах представляют собой уникальные хозяйствующие субъекты, призванные обеспечить устойчивое взаимодействие между высшими учебными заведениями и реальным сектором экономики.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,32;0,05;0,01;0,19;0,1; 0,1; 0,2)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №24

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,01; 0,12; 0,11; 0,03; 0,2; 0,14; 0,14; 0,25\}$.
2. Закодировать предложение: До Октябрьской революции на Кубани не было высших учебных заведений, которые готовили бы специалистов для сельского хозяйства.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,14;0,05;0,1;0,18;0,18; 0,2; 0,15)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №25

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=\{0,13; 0,06; 0,05; 0,4; 0,06; 0,1; 0,2\}$.
2. Закодировать предложение: За всю историю университета выпущено более 110 тыс. человек, среди которых более 9 тыс. – представители Азии, Африки и Латинской Америки.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,13;0,2;0,02;0,15;0,14; 0,1; 0,26)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №26

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=\{0,7; 0,18; 0,04; 0,01; 0,07\}$.
2. Закодировать предложение: Одна из основных форм международной деятельности университета в области образования – это подготовка иностранных студентов, магистров, аспирантов.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5\}$, $P=(0,13;0,4;0,07; 0,01; 0,39)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №27

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,011; 0,15; 0,15; 0,03; 0,2; 0,019; 0,04; 0,4\}$.
2. Закодировать предложение: Довузовская подготовка иностранных учащихся занимает особое место в образовательном процессе.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,15;0,03;0,4;0,01;0,11; 0,2; 0,1)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №28

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6\}$, $P=\{0,2; 0,14; 0,14; 0,1; 0,12; 0,3\}$.
2. Закодировать предложение: Научно-исследовательский институт прикладной и экспериментальной экологии Кубанского государственного аграрного университета создан в декабре 1995 г.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,23;0,1;0,04;0,03;0,1; 0,3; 0,2)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №29

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$, $P=\{0,18; 0,05; 0,17; 0,007; 0,093; 0,23; 0,17; 0,1\}$.
2. Закодировать предложение: Заказчиками института являются промышленные и сельскохозяйственные предприятия, муниципальные образования, государственные органы исполнительной власти регионального и федерального уровня.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, $P=(0,3;0,02;0,15;0,18;0,07; 0,18; 0,1)$
4. Для заданий 1-3 определить стоимость кодирования.

Вариант №30

Проведите кодирование следующих последовательностей символов.

1. Применить алгоритм Хаффмана: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$,
 $P=\{0,1; 0,17; 0,03; 0,03; 0,15; 0,04; 0,08; 0,4\}$.
2. Закодировать предложение: В настоящее время институт представляет собой крупный научно-исследовательский центр в области экологии.
3. Применить алгоритм Шеннона-Фано: $V=\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$,
 $P=(0,24; 0,17; 0,1; 0,3; 0,01; 0,1; 0,08)$
4. Для заданий 1-3 определить стоимость кодирования.