Федеральное агентство связи

Сибирский Государственный Университет Телекоммуникаций и Информатики

**Межрегиональный центр переподготовки специалистов**

# Контрольная работа

# По дисциплине: Теория информации

**Выполнил**: Бикшанов К. К.

**Группа**: ПБТ-49

**Вариант:9**

**Проверил**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Новосибирск, 2014 г

Формулировка задания

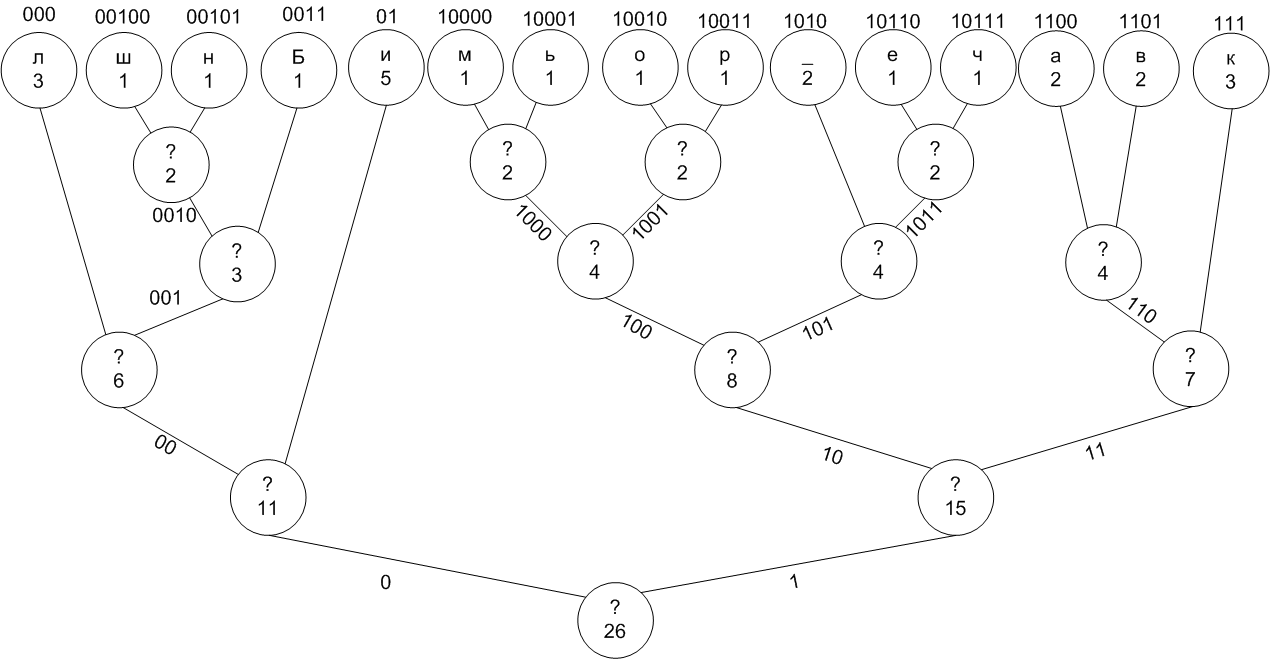
* Построить  код Хаффмана для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.
* Построить код Фано для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.
* Построить код Шеннона для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.
* Закодировать первые три буквы своего имени арифметическим кодом. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО.
* Закодировать последовательность из 10 букв ФИО адаптивным кодом Хаффмана (размер окна 6).
  1. Построить код Хаффмана для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.

Классический алгоритм Хаффмана на входе получает таблицу частот встречаемости символов в сообщении. Далее на основании этой таблицы строится дерево кодирования Хаффмана (Н-дерево). [[1]](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B4_%D0%A5%D0%B0%D1%84%D1%84%D0%BC%D0%B0%D0%BD%D0%B0#cite_note-1)

1. Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемое сообщение.
2. Выбираются два свободных узла дерева с наименьшими весами.
3. Создается их родитель с весом, равным их суммарному весу.
4. Родитель добавляется в список свободных узлов, а два его потомка удаляются из этого списка.
5. Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой — бит 0.
6. Шаги, начиная со второго, повторяются до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

Таблица 1. Таблица частот

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| k | Символы | Частоты вхождения | Вероятность состояния | Вычисление энтропии |
| 1 | и | 5 | 0,19 | 0,455226449 |
| 2 | к | 3 | 0,12 | 0,367067243 |
| 3 | л | 3 | 0,12 | 0,367067243 |
| 4 | а | 2 | 0,08 | 0,291508495 |
| 5 | в | 2 | 0,08 | 0,291508495 |
| 6 | пробел | 2 | 0,08 | 0,291508495 |
| 7 | Б | 1 | 0,04 | 0,185754248 |
| 8 | ш | 1 | 0,04 | 0,185754248 |
| 9 | н | 1 | 0,04 | 0,185754248 |
| 10 | о | 1 | 0,04 | 0,185754248 |
| 11 | р | 1 | 0,04 | 0,185754248 |
| 12 | м | 1 | 0,04 | 0,185754248 |
| 13 | ь | 1 | 0,04 | 0,185754248 |
| 14 | е | 1 | 0,04 | 0,185754248 |
| 15 | ч | 1 | 0,04 | 0,185754248 |
|  |  |  | Энтропия: | 3,735674648 |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| k | Символы | Частоты вхождения | Вероятность состояния | Кодовое слово | Длина кодового слова | Вычисления средней длины кодового слова |
| 1 | и | 5 | 0,19 | 1 | 2 | 0,38 |
| 2 | к | 3 | 0,12 | 111 | 3 | 0,36 |
| 3 | л | 3 | 0,12 | 000 | 3 | 0,36 |
| 4 | а | 2 | 0,08 | 1100 | 4 | 0,32 |
| 5 | в | 2 | 0,08 | 1101 | 4 | 0,32 |
| 6 | пробел | 2 | 0,08 | 1010 | 4 | 0,32 |
| 7 | Б | 1 | 0,04 | 0011 | 4 | 0,16 |
| 8 | ш | 1 | 0,04 | 00100 | 5 | 0,2 |
| 9 | н | 1 | 0,04 | 00101 | 5 | 0,2 |
| 10 | о | 1 | 0,04 | 10010 | 5 | 0,2 |
| 11 | р | 1 | 0,04 | 10011 | 5 | 0,2 |
| 12 | м | 1 | 0,04 | 10000 | 5 | 0,2 |
| 13 | ь | 1 | 0,04 | 10001 | 5 | 0,2 |
| 14 | е | 1 | 0,04 | 10110 | 5 | 0,2 |
| 15 | ч | 1 | 0,04 | 10111 | 5 | 0,2 |
|  |  |  |  |  | Сумма (Бит): | 3,82 |

Таблица 2 Кодовые слова Хаффмана. Подсчет средней длины кодового слова.

Результаты вычисления средней длины кодового слова построенного кода приведены в таблице 2.

* 1. Построить код Фано для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.

Построение кода фано выполнялось вручную при помощи табличного процессора Microsoft Excel. Использованы две вспомогательные ячейки, в которых записывались суммы вероятностей поделенного на две части списка алфавита. В третьей ячейке вычислялась разница между двумя ячейками сумм вероятностей. Результаты построения кода приведены в таблице 3.

Таблица 3. Построение кода Фано

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | Символы | Частоты вхождения | Вероянтность | Кодовое слово | | | | | Длина кодового слова | Вычисление средней длины кодового слова |
| 1 | и | 5 | 0,19 | 0 | 0 | 0 |  |  | 3 | 0,57 |
| 2 | к | 3 | 0,12 | 0 | 0 | 1 |  |  | 3 | 0,36 |
| 3 | л | 3 | 0,12 | 0 | 1 | 0 |  |  | 3 | 0,36 |
| 4 | а | 2 | 0,08 | 0 | 1 | 1 |  |  | 3 | 0,24 |
| 5 | в | 2 | 0,08 | 1 | 0 | 0 |  |  | 3 | 0,24 |
| 6 | пробел | 2 | 0,08 | 1 | 0 | 1 | 0 |  | 4 | 0,32 |
| 7 | Б | 1 | 0,04 | 1 | 0 | 1 | 1 | 1 | 5 | 0,2 |
| 8 | ш | 1 | 0,04 | 1 | 0 | 1 | 1 | 0 | 5 | 0,2 |
| 9 | н | 1 | 0,04 | 1 | 1 | 0 | 0 | 0 | 5 | 0,2 |
| 10 | о | 1 | 0,04 | 1 | 1 | 0 | 0 | 1 | 5 | 0,2 |
| 11 | р | 1 | 0,04 | 1 | 1 | 0 | 1 | 0 | 5 | 0,2 |
| 12 | м | 1 | 0,04 | 1 | 1 | 0 | 1 | 1 | 5 | 0,2 |
| 13 | ь | 1 | 0,04 | 1 | 1 | 1 | 0 |  | 4 | 0,16 |
| 14 | е | 1 | 0,04 | 1 | 1 | 1 | 1 | 0 | 5 | 0,2 |
| 15 | ч | 1 | 0,04 | 1 | 1 | 1 | 1 | 1 | 5 | 0,2 |
|  |  |  |  |  |  |  |  |  | Сумма(Бит) | 3,85 |

Средняя длина кодового слова равна 3,85 бит.Построить код Шеннона для набора букв ФИО. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО. Подсчитать среднюю длину кодового слова построенного кода.

Построение кода Шеннона проводилось вручную в табличном процессоре Microsoft Excel:

1. Вычисление кумулятивных вероятностей (Вычисляемое поле Q);
2. Вычисление длины кодового слова согласно выражению из конспекта лекций. Данные поля вероятность поделены на длину сообщения (26);
3. Кодовые слова получены переводом значения поля Q из десятичной системы счисления в двоичную, при помощи ресурса [1].
4. Вычисляем среднюю длину кодового слова как сумму последнего столбца таблицы.

Таблица 4. Построение кода Шеннона

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| m | Символы | Частоты вхождения | Вероянтность | Q | Длина кодового слова | Кодовое слово | Вычисление средней длины кодового слова |
| 0 |  |  |  | 0 |  |  |  |
| 1 | и | 5 | 0,19230769 | 0,192308 | 3 | 000 | 0,576923077 |
| 2 | к | 3 | 0,11538462 | 0,307692 | 4 | 0011 | 0,461538462 |
| 3 | л | 3 | 0,11538462 | 0,423077 | 4 | 0101 | 0,461538462 |
| 4 | а | 2 | 0,07692308 | 0,423077 | 4 | 0111 | 0,307692308 |
| 5 | в | 2 | 0,07692308 | 0,576923 | 4 | 1000 | 0,307692308 |
| 6 | пробел | 2 | 0,07692308 | 0,653846 | 4 | 1001 | 0,307692308 |
| 7 | Б | 1 | 0,03846154 | 0,692308 | 5 | 10101 | 0,192307692 |
| 8 | ш | 1 | 0,03846154 | 0,730769 | 5 | 10111 | 0,192307692 |
| 9 | н | 1 | 0,03846154 | 0,769231 | 5 | 11000 | 0,192307692 |
| 10 | о | 1 | 0,03846154 | 0,769231 | 5 | 11001 | 0,192307692 |
| 11 | р | 1 | 0,03846154 | 0,846154 | 5 | 11011 | 0,192307692 |
| 12 | м | 1 | 0,03846154 | 0,846154 | 5 | 11100 | 0,192307692 |
| 13 | ь | 1 | 0,03846154 | 0,923077 | 5 | 11101 | 0,192307692 |
| 14 | е | 1 | 0,03846154 | 0,961538 | 5 | 11110 | 0,192307692 |
| 15 | ч | 1 | 0,03846154 | 1 | 5 | 11111 | 0,192307692 |
|  |  |  |  |  |  | Сумма: | 4,153846154 |

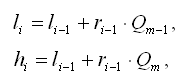
Закодировать первые три буквы своего имени арифметическим кодом. Для оценки вероятностей символов использовать частоты вхождения букв в ФИО.

Суть построения арифметического кода для первых трех букв сообщения сводится к вычислению границ интервала для третьего символа сообщения.

Воспользуемся таблицей 4.

Инициализируем начальные значения этих величин

и далее будем вычислять границы интервала, соответствующего кодируемой букве  по формулам:



где m -  порядковый номер кодируемой буквы в алфавите источника m=1,...,n,  а i– номер кодируемого символа в сообщении.

Получим границы интервала, соответствующего первому символу кодируемого сообщения “Б”(m=7):

Получим границы интервала, соответствующего второму символу кодируемого сообщения “и”(m=1):

Получим границы интервала, соответствующего третьему символу кодируемого сообщения “к”(m=2):

Кодом последовательности «Бик» будет двоичная запись любой точки из интервала [), например 0,655. Для однозначного декодирования возьмем   разрядов, получим код 101010.

Закодировать последовательность из 10 букв ФИО адаптивным кодом Хаффмана (размер окна 6).

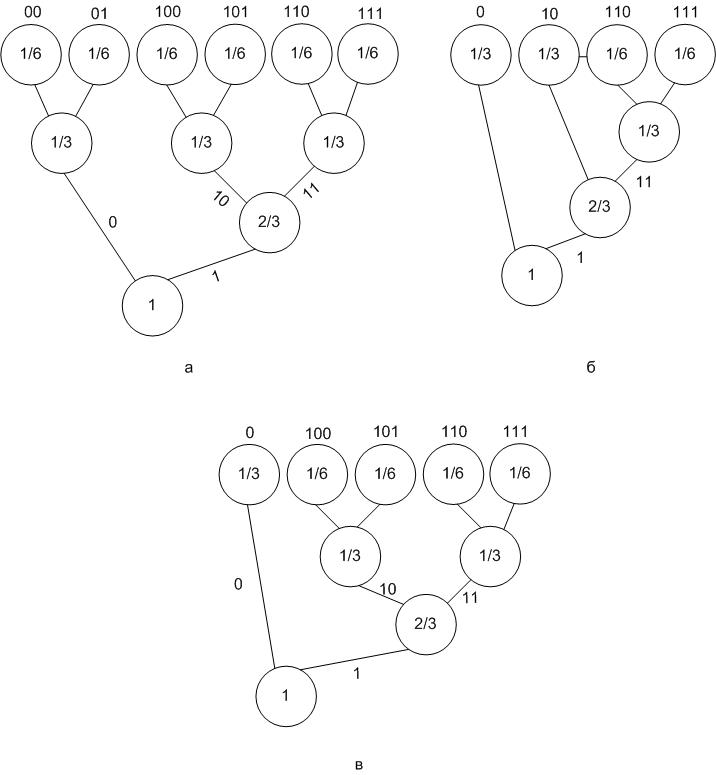
Последовательность из 10 букв ФИО:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| К | и | р | и | л | л | Б | и | к | ш |

Таблица 5. Кодирование адаптивным кодом Хаффмана

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **К** | и | р | и | л | л | Б | и | к | ш |  |  |  |  |  |  |  |  |
| Б | и | к | л | р | ш |  |  |  |  |  |  | Символы | Верояность | Построение | | | Слова |
| Инициализация окна символами алфавита; Построение кодов хаффмана для полученного окна; Считывание символа сообщения; 100 на выход кодера; | | | | | | | | | |  |  | Б | 1/6 | Рис 2а | | | 00 |
|  |  | и | 1/6 | 01 |
|  |  | к | 1/6 | 100 |
|  |  | л | 1/6 | 101 |
|  |  | р | 1/6 | 110 |
|  |  | ш | 1/6 | 111 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | **и** | р | и | л | л | б | и | к | ш |  |  |  |  |  |  |  |  |
| и | к | л | р | ш | к |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 100 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | к | 1/3 | Рис 2в | | | 0 |
|  |  | и | 1/6 | 100 |
|  |  | л | 1/6 | 101 |
|  |  | р | 1/6 | 110 |
|  |  | ш | 1/6 | 111 |
|  |  | Б | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | **р** | и | л | л | б | и | к | ш |  |  |  |  |  |  |  |  |
| к | л | р | ш | к | р |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 10 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | к | 1/3 | Рис 2б | | | 0 |
|  |  | р | 1/3 | 10 |
|  |  | л | 1/6 | 110 |
|  |  | ш | 1/6 | 111 |
|  |  | Б | 0 | - |
|  |  | и | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | **и** | л | л | б | и | к | ш |  |  |  |  |  |  |  |  |
| л | р | ш | к | р | и |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 110 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | р | 1/3 | Рис 2в | | | 00 |
|  |  | л | 1/6 | 01 |
|  |  | ш | 1/6 | 100 |
|  |  | к | 1/6 | 101 |
|  |  | и | 1/6 | 110 |
|  |  | Б | 0 | 111 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | **л** | л | б | и | к | ш |  |  |  |  |  |  |  |  |
| р | ш | к | р | и | л |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 100 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | р | 1/3 | Рис 2в | | | 0 |
|  |  | л | 1/6 | 100 |
|  |  | ш | 1/6 | 101 |
|  |  | к | 1/6 | 110 |
|  |  | и | 1/6 | 111 |
|  |  | Б | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | л | **л** | б | и | к | ш |  |  |  |  |  |  |  |  |
| ш | к | р | и | л | л |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 0 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | л | 1/3 | Рис 2в | | | 0 |
|  |  | р | 1/6 | 100 |
|  |  | ш | 1/6 | 101 |
|  |  | к | 1/6 | 110 |
|  |  | и | 1/6 | 111 |
|  |  | Б | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | л | л | **б** | и | к | ш |  |  |  |  |  |  |  |  |
| к | р | и | л | л | б |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 110 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | л | 1/3 | Рис 2в | | | 0 |
|  |  | р | 1/6 | 100 |
|  |  | к | 1/6 | 101 |
|  |  | и | 1/6 | 110 |
|  |  | Б | 1/6 | 111 |
|  |  | ш | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | л | л | б | **и** | к | ш |  |  |  |  |  |  |  |  |
| р | и | л | л | б | и |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 10 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | л | 1/3 | Рис 2б | | | 0 |
|  |  | и | 1/3 | 10 |
|  |  | р | 1/6 | 110 |
|  |  | Б | 1/6 | 111 |
|  |  | ш | 0 | - |
|  |  | к | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | л | л | б | и | **к** | ш |  |  |  |  |  |  |  |  |
| и | л | л | б | и | к |  |  |  |  |  |  |  |  |  |  |  |  |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 110 на выход кодера; | | | | | | | | | |  |  | Символы | Верояность | Построение | | | Слова |
|  |  | л | 1/3 | Рис 2б | | | 0 |
|  |  | и | 1/3 | 10 |
|  |  | к | 1/6 | 110 |
|  |  | Б | 1/6 | 111 |
|  |  | ш | 0 | - |
|  |  | р | 0 | - |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| К | и | р | и | л | л | б | и | к | **ш** |  |  |  |  |  |  |  |  |
| л | л | б | и | к | ш |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | Символы | Верояность | Построение | | | Слова |
| Модификация окна; Построение кодов хаффмана для полученного окна; Считывание очередного символа сообщения; 111 на выход кодера; Конец входной последовательности. | | | | | | | | | |  |  | л | 1/3 | Рис 2в | | | 0 |
|  |  | и | 1/6 | 100 |
|  |  | к | 1/6 | 101 |
|  |  | Б | 1/6 | 110 |
|  |  | ш | 1/6 | 111 |
|  |  | р | 0 | - |

Рисунок 2 Построение кода Хаффмана



Список использованных источников

# PLANETCALC [Электронный ресурс] : Перевод дробных чисел из одной системы счисления в другую : Онлайн калькуляторы. URL: http://planetcalc.ru/862/ (дата обращения: 05.11.2014).