|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ЦЕЛЬ РАБОТЫ**  Изучение принципа эффективного кодирования источника дискретных сообщений.  **ДОМАШНЕЕ ЗАДАНИЕ**   1. Изучить принцип эффективного кодирования источника дискретных сообщений (метод Хаффмена). 2. Осуществить кодирование каждого сообщения алфавита (см. таблицу 1), используя двоичный код:  * равномерный; * код Хаффмена, в соответствии с заданным вариантом.   Таблица 1 Вероятности появления сообщений алфавита   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | Вариант  Знак | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |  | 0,20 | 0,13 | 0,04 | 0,28 | 0,37 | 0,07 | 0,01 | |  | 0,05 | 0,17 | 0,17 | 0,04 | 0,13 | 0,09 | 0,03 | |  | 0,17 | 0,04 | 0,14 | 0,16 | 0,17 | 0,01 | 0,16 | |  | 0,24 | 0,26 | 0,26 | 0,02 | 0,07 | 0,23 | 0,13 | |  | 0,28 | 0,15 | 0,10 | 0,13 | 0,10 | 0,27 | 0,37 | |  | 0,02 | 0,07 | 0,11 | 0,07 | 0,07 | 0,16 | 0,20 | |  | 0,04 | 0,18 | 0,18 | 0,30 | 0,09 | 0,17 | 0,10 |  1. Определить значения и . 2. Рассчитать значения и .   Вариант для построения кода определяется по последней цифре пароля. При N>7 номер варианта равен N-7. Если N=0, то вариант 3.  **КРАТКАЯ ТЕОРИЯ**  К числу основных информационных характеристик источника сообщений относятся: количество информации в отдельных сообщениях, энтропия и производительность источника сообщений.  ***Количество информации.*** Единицей измерения количества информации является ***бит.*** Чем меньше вероятность появления того или иного сообщения, тем большее количество информации извлекается при его получении и наоборот. Если источник может выдавать одно из двух независимых сообщений и первое из них выдается с вероятностью , то интуитивно понятно, что сообщение не несет информации, ибо оно заранее известно получателю.  Количество информации, которое приходится на одно сообщение *,* определяется выражением  .  ***Среднее количество информации в сообщениях*** поступающих от источника*,* называется ***энтропией Н(А)***. Она определяется путем усреднения значений по всему объему алфавита и для источника без памяти [1] равна  . , (1)  где - объем алфавита А передаваемых сообщений.  **Э*нтропия*** *–* это мера неопределенности в поведении источника сообщений (ИС). Она равна нулю, если с вероятностью равной единице источником выдается всегда одно и то же сообщение (в этом случае неопределенность в поведении ИС отсутствует). Энтропия максимальна, если сообщения выдаваемые источником появляются независимо и с одинаковой вероятностью. В этом случае она равна .  Среднее количество информации, выдаваемое источником в единицу времени, называют производительностью источника и определяют по формуле:  ,  где *Т –* среднее время, отводимое на передачу одного сообщения.  Выражение для средней длительности сообщения имеет вид  ,  здесь - время передачи ***i****–*го сообщения заданного алфавита.  Сформулируем задачу статистического кодирования, которую частоприходится решать в технике документальной электросвязи. Каждое сообщение алфавита необходимо закодировать, используя двоичный код (символы 1 и 0). При этом выбранный код, **во-первых**, должен обеспечивать возможность однозначного декодирования, т.е. позволять по принятой последовательности символов “0” и “1” однозначно восстановить переданное сообщение. **Во-вторых**, на передачу сообщения в среднем должно быть затрачено минимальное число двоичных символов, что позволит передавать за единицу времени максимальное число сообщений. Полученный код можно считать оптимальным.  **Пример 1.** *Пусть . Некоторые возможные коды для букв алфавита А представлены в таблице 2.*  *Таблица 2*   |  |  |  |  | | --- | --- | --- | --- | |  | Код 1 | Код 2 | Код 3 | |  | 1 | 01 | 0 | |  | 0 | 10 | 10 | |  | 01 | 11 | 111 |   *Код 1 не является однозначно декодируемым кодом, так как комбинация является начальной частью комбинации . Для доказательства этого рассмотрим, например, двоичную последовательность 0101. Она может быть декодирована одним из сообщений:* ***.***  *Код 2 декодируется однозначно, поскольку все кодовые комбинации этого кода имеют равные длины и различны.*  *Код 3 также однозначно декодируемый, поскольку никакая его кодовая комбинация не является началом (префиксом) другой кодового слова.*  Код, обладающий тем свойством, что никакая более короткая комбинация не является началом другой более длинной комбинации кода, называют ***префиксным.*** Префиксные коды всегда однозначно декодируемы.  **Кодовое дерево для множества кодовых слов.**  **Рисунок 1**. Пример двоичного кодового дерева  Наглядное графическое изображение множества кодовых комбинаций можно получить, установив соответствие между сообщениями и концевыми узлами двоичного дерева. Пример двоичного кодового дерева изображен на рисунке 1.  Две ветви, идущие от корня дерева к узлам первого порядка, соответствуют выбору между “0” и “1” в качестве первого двоичного символа кодовой комбинации: левая ветвь соответствует “0”, а правая – “1”. Две ветви, идущие из узлов первого порядка, соответствуют второму символу кодовых комбинаций, левая означает “0”, а правая – “1” и т. д. Ясно, что последовательность символов каждой кодовой комбинации определяет необходимые правила продвижения от корня дерева до концевого узла, соответствующего рассматриваемому сообщению.  Формально кодовые комбинации могут быть приписаны также промежуточным узлам. Например, промежуточному узлу второго порядка на рисунке 1 можно приписать комбинацию 11, которая соответствует первым двум символам кодовых комбинаций, соответствующих концевым узлам, порождаемых этим узлом. Однако кодовые комбинации, соответствующие промежуточным узлам, не могут быть использованы для представления сообщений, так как в этом случае нарушается требование префиксности кода.  Требование, чтобы только концевые узлы сопоставлялись сообщениям, и обеспечивает условие при котором ни одна из кодовых комбинаций не совпадает с началом (префиксом) более длинной кодовой комбинации. Любой код, кодовые комбинации которого соответствуют различным концевым вершинам некоторого двоичного дерева, является префиксным, т. е. однозначно декодируемым.  На рисунке 2 изображены кодовые деревья, соответствующие кодам 2 и 3 (см. таблицу 2).  **Рисунок 2** Кодовые деревья для кодов 2 (а) и 3 (б)  Вернемся к рассмотренному ранее примеру 1. Пусть вероятность появления сообщения равна 0,2; сообщения - 0,3; – 0,5. Тогда среднее число двоичных символов , приходящихся на одно сообщение  для кода 3, составляет , а для кода 2 - , т. е. код 2 в среднем экономичнее кода 3.  Рассмотрим еще один код – код 4, который сообщению ставит в соответствие 10; *–* 11; – 0. Среднее число двоичных символов, приходящихся на одно сообщение этого кода: , т. е. код 4 экономичнее и кода 3, и кода 2. Спрашивается, можно ли предложить код, который будет экономичнее кода 4? Как построить самый экономичный код для данного источника сообщений? Ответы на эти и многие другие вопросы дает основная теорема кодирования, сформулированная и доказанная впервые К. Шенноном для дискретного канала связи без помех.  Известно, что нельзя закодировать сообщение двоичным кодом таким образом, чтобы средняя длина кодовых комбинаций была меньше, чем величина энтропии источника сообщений, т.е.  В [1] показано, что существует способ кодирования, при котором ,  где - сколь угодно малая величина.  **Метод Хаффмена.** Одним из часто используемых методов эффективного кодирования является так называемый метод Хаффмена. Пусть сообщения входного алфавита имеют соответственно вероятности их появления *.*  Тогда алгоритм кодирования Хаффмена состоит в следующем:  1. Сообщения располагаются в столбец в порядке убывания вероятности их появления. 2. Два самых маловероятных сообщения и объединяем в одно сообщение *,* которое имеет вероятность, равную сумме вероятностей сообщений *,* т. е. . В результате получим сообщения *,* вероятности которых *.* Полученные сообщения вновь располагаем в порядке убывания вероятностей. 3. Повторяем шаги 1 и 2 до тех пор, пока не получим единственное сообщение, вероятность которого равна 1. 4. Проводя линии, объединяющие сообщения и образующие последовательные подмножества, получаем дерево, в котором отдельные сообщения являются концевыми узлами. Соответствующие им кодовые комбинации можно определить, приписывая левым ветвям объединения символ “1”, а правым - “0”. *Впрочем, понятия “левые” и “правые” ветви в данном случае относительны.*  Так как в процессе кодирования сообщениям сопоставляются только концевые узлы, полученный код (код Хаффмена) является префиксным и следовательно всегда однозначно декодируемым.  ***В данной работе рекомендуется обозначать “1” ветвь идущую от узла в сторону сообщения с большей вероятностью появления.***  Построение кода Хаффмена для восьми сообщений, появляющихся с вероятностями 0,2; 0,2; 0,15; 0,13; 0,12; 0,1; 0,07; 0,03, иллюстрируется рисунками 3 и 4.  **Рисунок 3**  **Рисунок 4 Кодовое дерево**  Из таблицы 4 видно, что полученный код является неравномерным, причем сообщению с максимальной вероятностью появления соответствует минимальная длительность кодовой комбинации (2), а сообщению с минимальной вероятностью – максимальная (4).  Таблица 4   |  |  |  |  | | --- | --- | --- | --- | | № | Вероятность появления сообщения | Структура кодовой комбинации | Длительность кодовой комбинации | |  | 0,20 | 01 | 2 | |  | 0,20 | 111 | 3 | |  | 0,15 | 110 | 3 | |  | 0,13 | 101 | 3 | |  | 0,12 | 100 | 3 | |  | 0,10 | 000 | 3 | |  | 0,07 | 0011 | 4 | |  | 0,03 | 0010 | 4 |   Пусть переданная последовательность сообщений отображается двоичной последовательностью  0 1 1 1 1 1 1 0 1 0 1 0 0 1 0… (2)  Рассмотрим влияние одиночной ошибки во втором разряде принятой кодовой последовательности на результат декодирования. При этом получим  0 0 1 1 1 1 1 0 1 0 1 0 0 1 0… (3)  Полученная последовательность расшифровывается следующим образом  0 0 11 1 1 1 0 1 0 1 0 0 1 0… (4)  Из (4) видно, что искажение даже одного двоичного элемента (0  1) может привести к появлению ошибок в нескольких сообщениях (треку ошибок). Это является существенным недостатком рассмотренного метода кодирования.  Среднее число двоичных символов на одно сообщение алфавита объемом , для неравномерных двоичных кодов, определяется выражением  .  Эффективность неравномерных кодов оценивается следующими параметрами:  коэффициентом статистического сжатия, который характеризует уменьшение числа двоичных символов на знак при применении методов эффективного кодирования по сравнению с применением равномерного кода;  ,  где - средняя длина кодовой комбинации при равномерном кодировании,  – коэффициентом относительной эффективности, который показывает степень близости средней длины кодовой комбинации к теоретически возможному пределу .  .  **ОПИСАНИЕ ЛАБОРАТОРНОЙ РАБОТЫ**  Настоящая лабораторная работа выполняется на ЭВМ. Используемая программа эмулирует работу системы связи, использующей равномерное и неравномерное кодирование.  Программным путем эмулируются такие устройства как:  – блок ввода сообщения; – блок равномерного кодирования; – кодер Хаффмена; – блок определения длины кодовой комбинации; – блок ввода ошибки; – декодер Хаффмена; – блок определения ошибочных сообщений; – блок отображения.  Функциональная схема моделируемой системы связи представлена на рисунке 5.  Рисунок 5  При помощи данной программы возможно наблюдение таких процессов, как кодирование исходной последовательности сообщений (см. таблицу 1) кодером Хаффмена и блоком равномерного кодирования, внесение ошибки в закодированную последовательность сообщений, наблюдение трека ошибок и др.  Программа производит контроль правильности выполнения студентами домашнего задания.  Для выбора необходимого блока надо щелкнуть по его символьному обозначению левой клавишей мыши.  **Назначение эмулируемых блоков**  Блок ввода сообщений  Этот блок обеспечивает ввод исходной последовательности сообщений. Вводимая последовательность может состоять из сообщений алфавита таблицы 1 или из букв русского алфавита (набираемых на клавиатуре).  После окончания ввода сообщения следует нажать клавишу “Enter” или щелкнуть курсором мыши по клавише “OK”.  Блок равномерного кодирования  В этом блоке Вы можете наблюдать исходную последовательность сообщений, закодированную равномерным кодом.  Кодер Хаффмена  В этом блоке Вы можете наблюдать Вашу исходную последовательность сообщений, закодированную с использованием неравномерного префиксного кода Хаффмена.  Блок определения длины кодовой комбинации.  В этом блоке производится отображение длины Вашего закодированного сообщения и средней длины кодовой комбинации для случаев равномерного кодирования и кодирования кодом Хаффмена.  Блок ввода ошибки  Этот блок позволяет вносить одиночную или групповую ошибку в поток данных, что имитирует ошибки в реальных дискретных каналах и демонстрирует эффект трека ошибок.  Для внесения ошибки необходимо щелкнуть на нужных ярлычках, которые отображают номер искажаемого двоичного символа и произвести инверсию.  Декодер Хаффмена  Этот блок производит декодирование закодированного сообщения, при этом отображается восстановленное сообщение на выходе декодера.  Блок определения ошибочных сообщений  В этом блоке происходит сравнение последовательности двоичных элементов, переданных в линию связи, с принятой последовательностью. Ошибочные элементы отображаются красным цветом.  Блок отображения  В этом блоке происходит отображение принятого сообщения, а также выводятся характеристики качества передачи: количество ошибочно принятых двоичных элементов и сообщений.  **ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ**   1. **Проверка результатов расчетов домашнего задания.**   В окне “Контроль домашних расчетов” ввести номер варианта домашнего задания, двоичные последовательности для каждого сообщения источника, значения и . ***Внимание!*** ***При вводе расчетных значений следует обеспечить точность расчетов до второго знака после запятой.***   1. **Определение средней длины сообщения при передаче последовательностей, составленных из сообщений, имеющих разную вероятность появления.**   2.1 Открыть окно блока ввода сообщения щелчком мыши. 2.2 В раскрывающемся списке верхней строки выбрать “алфавит из домашнего задания”. 2.3 Составить три последовательности по 16 сообщений исходного алфавита (см. таблицу 1), полученные:   * + - чередованием двух наиболее вероятных сообщений (*например, для варианта 1* );     - повторением сообщения алфавита, вероятность появления которого равна или наиболее близка к , где – объем алфавита источника (*например, для варианта 1* );     - повторением сообщения алфавита, вероятность появления которого минимальна (*например, для варианта 1*).   2.4 Ввести соответствующую последовательность в нижнюю строку блока ввода сообщений. Для этого поместить курсор в нижнюю строку и последовательно ввести цифры, соответствующие номерам сообщений.  *Например: для последовательности , следует ввести* 4545…45  2.5 В блоке определения длины кодовой комбинации посмотреть для каждой последовательности сообщений среднюю длину кодовой комбинации на сообщение алфавита при равномерном и эффективном кодировании;   * 1. **Исследование влияния одиночной ошибки на результаты декодирования**   3.1 Составить и ввести произвольную комбинацию из 16 сообщений. 3.2 В окне блока ввода ошибки изменить один из элементов последовательности на противоположный. В данном окне необходимо указать номер элемента, в который нужно ввести ошибку и нажать кнопку “инвертировать”. В соответствующем разряде маски ошибки появится единица. 3.3 Переписать в отчет принятую двоичную последовательность из нижнего поля блока определения ошибочных сообщений и декодировать ее в соответствии с кодовыми комбинациями Вашего варианта. Определить количество неверно принятых сообщений. 3.4 Проверить правильность декодирования, используя данные из блока отображения. 3.5 Установить в блоке ввода сообщения “русский алфавит”. 3.6 Ввести последовательность слов, состоящую из букв русского алфавита, где . 3.7 Посчитать и сравнить количество двоичных символов необходимых для передачи введенного текста при кодировании равномерным кодом и кодом Хаффмена. 3.8 Ввести ошибку в единичный элемент, соответствующий k-ой букве, где k номер варианта домашнего задания. 3.9 Посмотреть как расшифровывается последовательность, содержащая ошибку. Сделать выводы.  **КОНТРОЛЬНЫЕ ВОПРОСЫ**   * 1. Принцип формирования кодовых комбинаций при кодировании методом Хаффмена.   2. Как рассчитывается средняя длина кодовой комбинации кода Хаффмена и каково ее минимальное значение?   3. В чем состоит свойство префиксности эффективных кодов?   4. Количественные показатели эффективности неравномерного кодирования.   5. Принцип декодирования последовательности префиксного кода.   6. Принципы возникновения трека ошибок при декодировании последовательности кодовых комбинаций префиксного кода.   **СОДЕРЖАНИЕ ОТЧЕТА**   * 1. Результаты выполнения домашнего задания.   2. Результаты выполнения моделирования.   3. Выводы по работе.   4. Ответы на контрольные вопросы.   **СПИСОК ЛИТЕРАТУРЫ**   * 1. Тема 4. Эффективное кодирование.   2. Передача дискретных сообщений: Учебник для вузов/ В.П. Шувалов, Н.В. Захарченко, В.О. Шварцман и др.; под. ред. В.П. Шувалова. -М.: Радио и связь,-1990.- 464с.   3. Финк Л.М. Теория передачи дискретных сообщений. - М: Сов. радио, 1970.   4. Флястер И.И. Эффективное кодирование на примере кода Хаффмена //Методические указания по выполнению лабораторных работ.- М.: МТУСИ, 1991 г. |