**Лабораторная работа № 6**

**ТЕХНОЛОГИЯ СЖАТИЯ ДАННЫХ**

                            1. **Цель работы**

        Освоение технологии сжатия данных.

2. **Общие сведения**

**2.1. Методы сжатия данных**

Методы сжатия находят широкое применение при хранении и передаче текстовых, графических, аудио и видео данных. Объектами сжатия могут быть файлы, папки и диски. Методы сжатия информации делятся на методы сжатия без потерь (обратимое сжатие) информации и с потерями информации (необратимое сжатие).

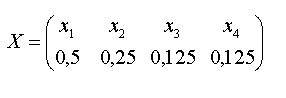
Методы сжатия  без потерь информации делятся на методы сжатия источников информации без памяти (метод Хаффмена, арифметическое сжатие, и др.) и методы сжатия источников информации с памятью. К последним относятся алгоритмы Лемпеля-Зива (LZ), Лемпеля-Зива-Велча (LZW) и др. На основе названных и других методов сжатия работают различные программы сжатия данных (архиваторы). К форматам сжатия без потери информации относятся: .ZIP, .ARJ, .RAR, .LZH, .LH, .CAB и др. (сжатие любых типов данных);  .GIF, .TIF, .PCX и  др. (сжатие графических данных).

Методы сжатия с потерей информации используются для архивации графических, аудио и видео данных, которые в несжатом виде требуют огромных объемов дискового пространства. С этими методами связано понятие качества сжатия, понимаемое как степень соответствия исходного и воспроизведенного изображения. Оценки качества сжатия обычно субъективны. (см. темы квантование сигнала по уровню и дискретизация сигнала во времени). Форматы сжатия с потерей информации:  .JPG для графических данных;  .MPG для видеоданных; .MP3 для звуковых данных.

Исторически первыми кодами сжатия информации без потерь информации являются, ставшие классикой, коды Шеннона-Фано и   Хаффмена. Названные коды исследуются в настоящей лабораторной работе.

**2.1. Методика построения кода Шеннона-Фано**

Проиллюстрируем методику построения кода Шеннона-Фано на примере источника сообщений, который описывается следующим рядом распределения  вероятностей

****

       Буквы алфа­вита источника информации выписываются в столбец в порядке убыва­ния их вероятностей (см. табл..1). Столбец разбивается на две подгруппы с равны­ми (по возможности) суммарными вероятностями. Каждая подгруппа, которая содержит более одной буквы, в свою очередь разбивается та­ким же образом на две подгруппы и т.д. Описанный процесс продолжа­ется до тех пор, пока во всех подгруппах очередного шага разбиения не останется по одной букве.

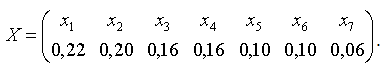
Таблица 1. Примерпостроения кода Шеннона–Фано

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | P(xi) | Номера деления       на группы | | | Символы  кода | | | Длина |
|  | | |
| 1 | 2 | 3 | 1 | 2 | 3 |
| x1 | 0,5 | 0 |  |  | 0 |  |  | 1 |
| x2 | 0,25 | 1 | 0 | 1 | 0 |  | 2 |
| x3 | 0,125 | 1 | 0 | 1 | 1 | 0 | 3 |
| x4 | 0,125 | 1 | 1 | 1 | 1 | 3 |

       Код формируется следующим образом. Всем верхним подгруппам каж­дого шага разбиения приписывается символ «0», а нижним — символ «1» (можно и наоборот). Длина кодовой комбинации буквы определяется числом шагов разбиения, в которых эта буква участвовала. Кодовая комбинация буквы формиру­ется слева неправо путем записи символов 0 или 1 в зависимости от того, в какую подгруппу (верхнюю или нижнюю) попала данная буква в соответствующем шаге.

**2.3. Методика построения кода Хаффмена**

       Код Хаффмена можно строить таблично, подобно коду Шениона-Фано, или графически. Графическое построение более наглядно. Рассмотрим методику графического построения кода Хаффмена для источника сообщений, который описывается следующим рядом распределения вероятностей



       Выпишем буквы алфавита источника в стол­бец в порядке убывания их вероятностей. Две буквы с наименьшими веро­ятностями  объединяем так, как это показано на рис. 1 в одну вспомогательную букву, которой припишем суммарную вероятность объединяемых букв. Среди оставшихся букв, включая вспомогательную, вновь находим две буквы с наименьшими вероятностями и повторяем описанную выше процедуру до тех пор, пока не будет получена  един­ственная вспомогательная буква с вероятностью, равной единице.

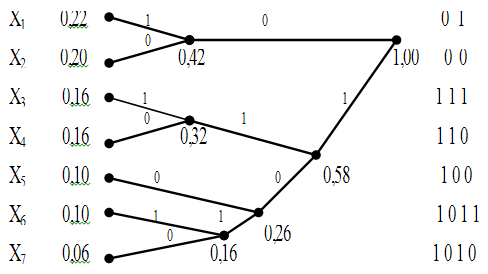


 Рис.1. Кодовое дерево Хаффмена

       Далее, всем ребрам полученного таким образом кодового дерева приписываются символы «0» или «1» по следующему правилу: из двух ребер, выходящих влево из одной вершины, ребру, соединенному с вершиной, имеющей большую (или равную) вероятность, приписывается 1, а ребру, соединенному с вершиной, имеющей меньшую вероятность, — 0. Ко­довая комбинация буквы составляется, начиная с вершины с вероят­ностью "единица", путем последовательной записи символов, находящихся на ребрах цепи, соединяющей эту вершину с   соответствующей   буквой алфавита.

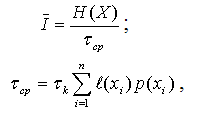
       В архиваторах, использующих метод Хаффмена, могут применяться алгоритмы, основанные либо на применении готовых частотных таблиц, либо такие таблицы строятся в процессе статистического анализа содержимого сжимаемого файла.

**2.4. Задачи лабораторной  работы**

1, Постройте код Шеннона-Фано и код Хаффмена для источника сообщений, который описывается случайной величиной Х из 2 задачи  лабораторной работы № 5.

2. Вычислите скорость передачи информации в канале связи, задан­ном в задаче 5 лабораторной работы № 5 , при условии, что сообщения источника   закодированы в коде Шеннона-Фано.

Указания. Скорость передачи информации в бинарном канале без шума при использовании неравномерного кода определяет­ся по формуле



где Описание: Описание: E:\ЭРУД\ЭРУД_ОинфТ\Практика\content\lb6\img141.gif− количество элементарных кодовых посылок, необходи­мых для передачи буквы Описание: Описание: E:\ЭРУД\ЭРУД_ОинфТ\Практика\content\lb6\img143.gif− источника дискретных сообщений.

Сравните результаты задачи 5 лабораторной работы № 5,  и задачи 2  этой лабораторной работы; сделайте необходимые выводы.

**3. Порядок выполнения работы**

        3.1. Ознакомьтесь  с материалом, изложенным в пункте 2, этой  работы.

        3.2. Постройте коды Шеннона-Фано и Хаффмена по исходным данным, полученным в лабораторной работе  № 5.

 3.3. Вычислите скорость передачи информации в канале связи при условии, что сообщения источника  закодированы в коде Шеннона-Фано.

 3.4.  Проанализируйте полученные результаты и сделайте необходимые выводы.

         3.5. Оформите отчет и защитите лабораторную работу.

        3.6. Выключите компьютер и приведите в порядок рабочее  место.

**4. Содержание отчета**

      4.1. Цель работы.

 4.2. Коды Шеннона-Фано и Хаффмена.

 4.3. Расчет скорости передачи информации в канале связи  при использовании кода Шеннона-Фано.

      4.4. Анализ полученных результатов.

      4.5. Выводы по работе.

**5. Контрольные вопросы**

   1. Охарактеризуйте методы сжатия информации.

        2. В каких случаях применяются методы сжатия с потерей информации?

   3. В чем состоят достоинства кода  Хаффмена?

  4. За счет чего при использовании кода  Шеннона-Фано обеспечивается повышение скорости передачи информации в канале связи?

        5. Какие алгоритмы применяются в архиваторах, использующих метод Хаффмена?