МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет информационных технологий и управления

Кафедра информационных технологий автоматизированных систем

**ОТЧЕТ**

По лабораторной работе №6

«Технология сжатия данных»

|  |  |  |
| --- | --- | --- |
| Выполнил |  | Проверил: |
| ст. гр. 328501 |  | Ярмолик В. И. |
| Суворов В. В. |  |  |

Минск 2023

**1 ЦЕЛЬ РАБОТЫ**

Изучение и 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).

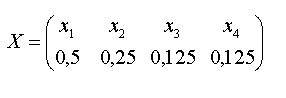
Методы сжатия без потерь информации делятся на методы сжатия источников информации без памяти (метод Хаффмена, арифметическое сжатие, и др.) и методы сжатия источников информации с памятью. К последним относятся алгоритмы Лемпеля-Зива (*LZ*), Лемпеля-Зива-Велча (*LZW*) и др. На основе названных и других 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.

**2.2 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 н а две подгруппы и т.д. Описанный процесс продолжа­ется до тех пор, пока во всех подгруппах очередного шага разбиения не останется по одной букве.

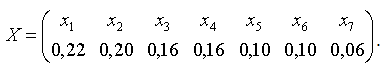
Таблица 1 – Примерпостроения кода Шеннона-Фано

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | *P(xi)* | Группа | | | Символ 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.

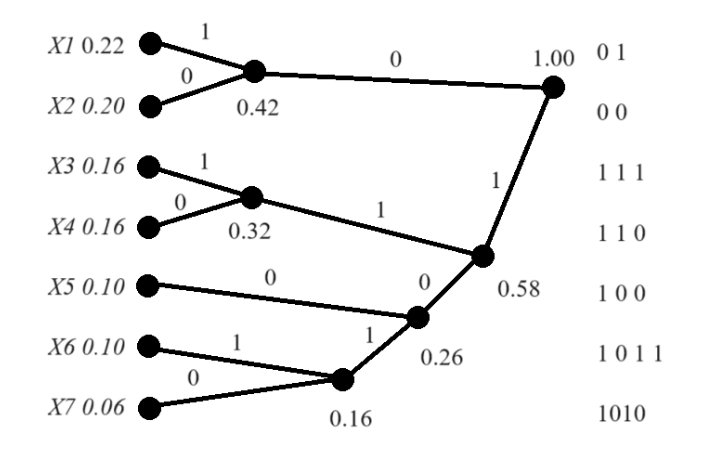


Рисунок 1 - Кодовое дерево Хаффмена.

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.

**3 ПРАКТИЧЕСКОЕ ЗАДАНИЕ**

Даны вероятности х1 = 0,21; х2 = 0,17; х3 = 0,09; х4 = 0,26; х5 = 0,031; х6 = 0,14; х7 = 0,042; х8 = 0,057. Вычислить код Шеннона-Фано и код Хаффмана.

**3.2 Код Шеннона-Фано**

Таблица 2 - Построения кода Шеннона-Фано.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Х4** | **0,26** | **1** | **1** |  |  |  | **11** |
| **Х1** | **0,21** | **1** | **0** |  |  |  | **10** |
| **Х2** | **0,17** | **0** | **1** | **1** |  |  | **011** |
| **Х6** | **0,14** | **0** | **1** | **0** |  |  | **010** |
| **Х3** | **0,09** | **0** | **0** | **1** |  |  | **001** |
| **Х8** | **0,057** | **0** | **0** | **0** | **1** |  | **0001** |
| **Х7** | **0,042** | **0** | **0** | **0** | **0** | **1** | **00001** |
| **Х5** | **0,031** | **0** | **0** | **0** | **0** | **0** | **00000** |

Следовательно, х1 = 10; х2 = 011; х3 = 001; х4 = 11; х5 = 00000; х6 = 010; х7 = 00001; х8 = 0001. Т.е. код будет выглядеть как 10 011 001 11 00000 010 00001 0001.

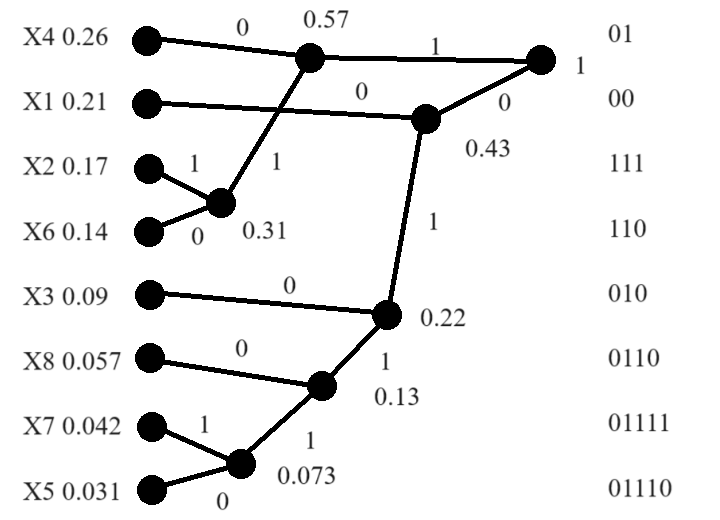
**3.3 Код Хаффмана**

Рисунок 2 – Построение кода Хаффмана

Следовательно, х1 = 00; х2 = 111; х3 = 01

0; х4 = 10; х5 = 01110; х6 = 110; х7 = 01111; х8 = 0110. Т.е. код будет выглядеть как 00 111 010 10 01110 110 01111 0110.

**4 АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ**

Из полученных результатов можно заметить, что сжатие данных кодом Шеннона-Фано и кодом Хаффмана достаточно сильно отличаются.

**5 ВЫВОД**

Мы доказали, что сжатие данных кодом Шеннона-Фано и кодом Хаффмана очень сильно различаются между собой. Из этого следует, что разные методы сжатия данных можно применять в разных сферах. Вообще, методы сжатия находят широкое применение при хранении и передаче текстовых, графических, аудио и видео данных. Объектами сжатия могут быть файлы, папки и диски.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Х2** | **0,17** | **1** | **1** | **1** |  |  | **111** |
| **Х6** | **0,14** | **1** | **1** | **0** |  |  | **110** |
| **Х4** | **0,26** | **1** | **0** |  |  |  | **10** |
| **Х7** | **0,042** | **0** | **1** | **1** | **1** | **1** | **01111** |
| **Х5** | **0,031** | **0** | **1** | **1** | **1** | **0** | **01110** |
| **Х8** | **0,057** | **0** | **1** | **1** | **0** |  | **0110** |
| **Х3** | **0,09** | **0** | **1** | **0** |  |  | **010** |
| **Х1** | **0,21** | **0** | **0** |  |  |  | **00** |

.