|  |
| --- |
| https://www.mirea.ru/bitrix/templates/unlimtech/images/logo.png |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** |

|  |  |
| --- | --- |
| Отчет по выполнению практического задания №7  По дисциплине «Структуры и алгоритмы обработки данных» | |
| **Тема:**  Кодирование и сжатие данных методами без потерь | |
|  | |
| Выполнил студент | Сидоров С.Д. |
| группа | ИКБО-20-21 |

**Отчёт**

Цель: получение практических навыков и знаний по выполнению сжатия данных рассматриваемыми методами

**Задание 1.** Применение алгоритма группового сжатия текста.

Постановка задачи: сжать текст, используя метод RLE.

Выполнение:

1. Сжатие методом RLE происходит за счёт выделения цепочек символов в тексте. Каждая серия символов записывается, в виде «i, “s”», где i - число повторений символа, а “s” сам символ. Число повторений в одной группе может быть разбито на подгруппы по 127 повторений в одной, в случае если используется «*signed char*». Также для работы с последовательностями, где встречаются длинные наборы неповторяющихся символов были введены отрицательные маркеры количества повторений. В случае, если в записи «i, “s”» i - отрицательное, то это означает, что следующие |i| символов представляют собой набор неповторяющихся символов.
2. Для первого примера возьмём строку “AABBBCCCDDDDEFFFFFFGGGHHHHH”.

Данная строка подразбивается на группы: 2A, 3B, 3C, 4D, 1E, 6F, 3G, 5H. Конечный результат сжатия будет представлять собой строку: “2A3B3C4D1E6F3G5H”.

Рассчитаем коэфицент сжатия:

Длина исходного текста: 27 символов.

Длина сжатого текста: 16 символов.

Коэфицент сжатия в данном случае равен: 16/27 ≈ 0.6

1. Для второго примера возьмём похожую строку, но разбавим её последовательностями из неповторяющихся подряд символов. Строк - DEFFHTYRUFFGGGHHH”.

Произведём сжатие без применения разделения текста. Данная строка подразбивается на группы: 1D, 1E, 2F, 1H, 1T, 1Y, 1R, 1U, 2F, 3G, 3H . Конечный результат сжатия будет представлять собой строку: “1D1E2F1H1T1Y1R1U2F3G3H”.

Рассчитаем коэфицент сжатия:

Длина исходного текста: 17 символов.

Длина сжатого текста: 22 символа.

Коэфицент сжатия в данном случае равен: 22/17≈ 1.294117647. Коэфицент сжатия больше единицы следовательно сжатый текст занимает больше места, чем оригинальный.

Теперь применим метод разделения текста для RLE. В таком случае последовательности неповторяющихся подряд символов будут закодированного с помощью отрицательных чисел.

Данная строка подразбивается на группы: -2DE, 2F, -5HTYRU, 2F, 3G, 3H . Конечный результат сжатия будет представлять собой строку: “-2DE2F-5HTYRU2F3G3H”.

Рассчитаем коэфицент сжатия:

Длина исходного текста: 17 символов.

Длина сжатого текста: 19 символа.

Коэфицент сжатия в данном случае равен: 19/17≈ 1.117647059. Коэфицент сжатия больше единицы следовательно сжатый текст занимает больше места, чем оригинальный, однако если сравнить коэфиценты, заметно что при применении данного метода сжатия получается выгоднее, следовательно, данный метод улучшает сжатие в больших текстах содержащих, как длинные последовательности одинаковых символов, так и последовательности символов неповторяющихся подряд.

**Задание 2.**

Для метода Лемпеля - Зива LZ77 для сжатия двоичного кода.

Вариант 1 : 0001010010101001101

|  |  |
| --- | --- |
| Исходный текст | 0001010010101001101 |
| LZ-код | 0.00.10.110.111.001.0110.0111 |
| R | 2 3 4 |
| Вводимые коды | - 10 11 100 101 110 111 1000 |

LZ- сжатый текст, в данном случае из-за небольшого размера исходного текста, в следствие сжатия размер текста не уменьшился.

Для метода Лемпеля –Зива LZ78 для сжатия текста.

Вариант 1 : кукуркукурекурекун

|  |  |  |
| --- | --- | --- |
| Содержимое словаря | Содержимое строки | Код |
| к | к | «0,к» ( 1 ) |
| к, у | у | «0,у» ( 2 ) |
| к, у, ку | ку | «1,у» ( 3 ) |
| к, у, ку, р | р | «0,р» ( 4 ) |
| к, у, ку, р, кук | кук | «3, к» ( 5 ) |
| к, у, ку, р, кук, ур | ур | «2, р» ( 6 ) |
| к, у, ку, р, кук, ур, е | е | «0,е» ( 7 ) |
| к, у, ку, р, кук, ур, е, кур | кур | «3, р» ( 8 ) |
| к, у, ку, р, кук, ур, е, кур, ек | ек | «7,к»( 9 ) |
| к, у, ку, р, кук, ур, е, кур, ек, ун | ун | «2, н» ( 10 ) |

Сжатый текст: 0к0у1у0р3к2р0е3р7к2н

Для преобразования данного текста в изначальный необходимо пройтись последовательно с начала сжатой строки, записывая символы и их сочетания в словарь, что бы в дальнейшем использовать их при подстановке.

**Задание 3.**

Метод Шенона - Фано

Вариант 1: Ана, дэус, рики, паки, Дормы кормы констунтаки, Дэус дэус канадэус – бац! (Длина - 73 )

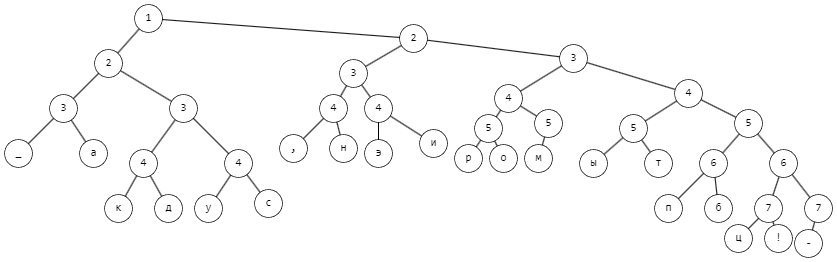
1. Таблица:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Символ** | **Кол-во** | **1-я цифра** | **2-я цифра** | **3-я цифра** | **4-я цифра** | **5-я цифра** | **6-я цифра** | **7-я цифра** | **Код** | **Кол-во бит** |
| \_ | 11 | 0 | 0 | 0 |  |  |  |  | 000 | 33 |
| а | 7 | 0 | 0 | 1 |  |  |  |  | 001 | 21 |
| к | 6 | 0 | 1 | 0 | 0 |  |  |  | 0100 | 24 |
| д | 5 | 0 | 1 | 0 | 1 |  |  |  | 0101 | 20 |
| у | 5 | 0 | 1 | 1 | 0 |  |  |  | 0110 | 20 |
| с | 5 | 0 | 1 | 1 | 1 |  |  |  | 0111 | 20 |
| , | 5 | 1 | 0 | 0 | 0 |  |  |  | 1000 | 20 |
| н | 4 | 1 | 0 | 0 | 1 |  |  |  | 1001 | 16 |
| э | 4 | 1 | 0 | 1 | 0 |  |  |  | 1010 | 16 |
| и | 4 | 1 | 0 | 1 | 1 |  |  |  | 1011 | 16 |
| р | 3 | 1 | 1 | 0 | 0 | 0 |  |  | 11000 | 15 |
| о | 3 | 1 | 1 | 0 | 0 | 1 |  |  | 11001 | 15 |
| м | 2 | 1 | 1 | 0 | 1 | 0 |  |  | 11010 | 10 |
| ы | 2 | 1 | 1 | 1 | 0 | 0 |  |  | 11100 | 10 |
| т | 2 | 1 | 1 | 1 | 0 | 1 |  |  | 11101 | 10 |
| п | 1 | 1 | 1 | 1 | 1 | 0 | 0 |  | 111100 | 6 |
| б | 1 | 1 | 1 | 1 | 1 | 0 | 1 |  | 111101 | 6 |
| ц | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1111100 | 7 |
| ! | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1111101 | 7 |
| - | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1111111 | 7 |

Объем незакодированной фразы – 73\*8 бит = 584 бит.

Объем закодированной фразы –289 бит.

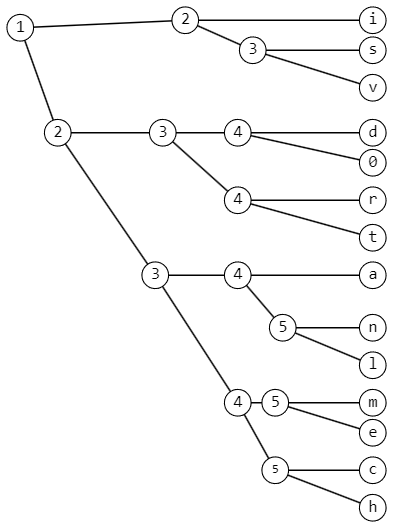
Коэффициент сжатия: 289/584 =0.5



Метод Хаффмана:

Строка: “sidorovstanislavdmitrievich”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Алфавит | i | s | v | d | o | r | t | a |
| Кол-во | 5 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| Веряотн. | 0.18 | 0.11 | 0.11 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 |
|  |  |  |  |  |  |  |  |  |
| Алфавит | n | l | m | e | c | h |  |  |
| Кол-во | 1 | 1 | 1 | 1 | 1 | 1 |  |  |
| Веряотн. | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |  |  |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | i | s | v | d | o | r | t |
| Код | 00 | 010 | 011 | 1000 | 1001 | 1010 | 1011 |
|  |  |  |  |  |  |  |  |
| Символ | a | n | l | m | r | c | h |
| Код | 1100 | 11010 | 11011 | 11100 | 11101 | 11110 | 11111 |

Объем незакодированной фразы – 27\*8 бит = 216 бит.

Объем закодированной фразы –110 бит.

Коэффициент сжатия: 110/216=0.5

Средняя длина: 4,074

**Задание 4:**

Код программы представлен в файлах:

Node.pdf, frequencies.pdf, tree.pdf, source.pdf.

Исходный текстовый файл занимает 53\*8 = 424 байта.

После сжатия методом Хаффмана 211 байт.

После программного сжатия стало 53\*8 = 424 байта.

Выводы:

В результате выполнения данной работы были освоены навыки работы с алгоритмами сжатия Лемпеля-Зива LZ77 и LZ78, Шеннона-Фано и Хаффмана. Также мной был получен практический опыт использования и реализации различных алгоритмов сжатия текста и других данных.

**Список литературы:**

⦁ Лекции по структурам и алгоритмам обработки данных Рысин М.Л.

⦁ Методическое пособие по выполнению задания 1(битовые операции)