МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное

учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

|  |
| --- |
| Институт информационных технологий |
| Кафедра математического и программного обеспечения |
| Теория информации |
|  |

ЛАБОРАТОРНАЯ РАБОТА №4

«Исследование свойств форматов сжатия графических и текстовых данных»

|  |  |
| --- | --- |
| Исполнитель: | студент  группы 1ПИб-02-3оп-23  Богданов  Ренат Алексеевич |
| Руководитель: | Ганичева Оксана Георгиевна |
| Оценка: |  |
| Подпись: |  |

2024, Череповец

Задание

Задание 1

1. Откройте графический редактор Paint. Загрузите в него многоцветный рисунок (например, С:\Windows\Облака.bmp).
2. Определите размер рисунка в пикселах. Оцените теоретический размер рисунка в 24-разрядной палитре ( 3 байта на точку) по формуле:

*S= M·N·3* , где

S - размер файла с рисунком (байт)

M - ширина рисунка (точек)

N - высота рисунка (точек).

4. Сохраните рисунок под именем Рисунок\_1 , назначив тип файла : 24-разрядный рисунок (.bmp). (в заранее созданную папку с именем Мои\_Рисунки)

5. Повторно сохраните рисунок , с именем Рисунок\_2 , назначив тип файла .gif. (в заранее созданную папку с именем Мои\_Рисунки)

При сохранении произойдет потеря определенной части графической информации.

6. Восстановите рисунок, загрузив его из ранее сохраненного файла Рисунок\_2.bmp, и вновь сохраните его под именем Рисунок\_3, назначив тип файла .jpeg. (в заранее созданную папку с именем Мои\_Рисунки)

7. Запустите программу Проводник. Откройте папку Мои\_Рисунки в режиме Таблица. Определите размеры файлов Рисунок\_1.bmp, Рисунок\_2.gif Рисунок\_3.jpeg. Определите коэффициенты сжатия файлов К, взяв отношения размеров файлов к теоретической величине, полученной расчетным путем в п.3 Результаты занесите в таблицу:

|  |  |  |
| --- | --- | --- |
| Формат файла | Размер файла (Кбайт) | Степень сжатия (%) |
| 24-разрядный рисунок .bmp |  |  |
| .gif |  |  |
| .jpeg |  |  |

8. Сделайте вывод о степени сжатия данных в разных форматах

9. В графическом редакторе Paint дайте команду создания нового документа. Убедитесь в том, что полотно имеет размер 640х480. Если это не так, измените его размер.

10. В качестве инструмента выберите Кисть. Задайте максимальный размер кисти. Поочередно используя 8-10 разных красок, грубо закрасьте полотно.

11. Сохраните рисунок под именем Test\_1 в формате 24-разрядный рисунок .bmp.

12. Сохраните рисунок под именем Test\_2 в формате .gif.

13. Восстановите рисунок из файла Test\_1.bmp

14. Сохраните рисунок под именем Test\_3 в формате .jpeg.

15. С помощью программы проводник определите размеры сохраненных файлов и заполните таблицу.

|  |  |  |
| --- | --- | --- |
| Формат файла | Размер файла (Кбайт) | Степень сжатия (%) |
| 24-разрядный рисунок .bmp |  |  |
| .gif |  |  |
| .jpeg |  |  |

16. Сделайте вывод о степени сжатия файлов разных форматах.

Какой формат графических данных из рассмотренных в работе наилучшим образом подходит для передачи цветного фотографического материала по каналам электронных сетей?

Какой формат графических файлов данных целесообразно использовать для передачи черно-белого фотографического материала по каналам связи?

Какой формат наиболее подходит для передачи рисунков, имеющих малое количество цветовых оттенков (до 256).

Задание 2. Выполнить сжатие информации методом RLE

Выполнить вручную кодирование сообщения методом RLE. В качестве исходной фразы взять текст из табл. 1. С помощью таблицы CP-1251 перевести символы заданной фразы в десятичные числа, а затем десятичные числа перевести в двоичные. Выполнить сжатие информации, вычислить контрольные суммы и коэффициент сжатия.

Таблица 1

|  |  |
| --- | --- |
| Вар | Текст |
| 19 | Урааааааааааааа в атаку |

Задание 3. Выполнить сжатие информации методом Шеннона-Фано

Используя фразу из табл. 1, построить код Шеннона-Фано и определить коэффициент и степень сжатия этим методом

Задание 4. Выполнить сжатие информации методом Хаффмана

Используя фразу из табл. 1, построить кодовое дерево и определить коэффициент и степень сжатия методом Хаффмана

Задание 5. Сделать общий вывод о степени сжатия исходного текста (фраза из табл.1) каждым методом.

Ход работы

Задание 1

Откроем изображение sample.bpm разрешением 1920x1280 в графическом редакторе Paint. Изображение имеет 24 разрядную палитру. По формуле (1.1) находим теоретический размер изображения.

|  |  |  |
| --- | --- | --- |
|  |  | (1.1) |

Так, S = 1920 \* 1280 \* 3 байт = 7372800 байт = 7200 КБ.

Сохраним изображение в разных форматах и занесём данные в таблицу 2. Определим коэффициенты сжатия по формуле (1.2).

|  |  |  |
| --- | --- | --- |
|  |  | (1.2) |

где – исходный объём информации,

– объём сжатой информации.

Таблица 2

|  |  |  |
| --- | --- | --- |
| Формат файла | Размер файла (Кбайт) | Степень сжатия (%) |
| 24-разрядный рисунок .bmp | 7201 | 100 |
| .gif | 1194 | 603 |
| .jpeg | 1113 | 647 |

Таким образом, исходя из таблицы 1, можно сделать вывод, что формат .bmp не использует сжатие и предназначается для хранения необработанных изображений. Формат .jpeg обеспечивает наибольшее сжатие файла.

В графическом редакторе Paint создадим пустое изображение с разрешением 640\*480 и с помощью инструмента “Кисть” разрисуем его.

По формуле (1.1) подсчитаем теоретический размер: S =640 \* 480 \* 3 байт = 921600 байт = 900 КБ. Также сохраним изображение в разных форматах и занесём данные в таблицу 3.

Таблица 3

|  |  |  |
| --- | --- | --- |
| Формат файла | Размер файла (Кбайт) | Степень сжатия (%) |
| 24-разрядный рисунок .bmp | 901 | 100 |
| .gif | 125 | 720 |
| .jpeg | 184 | 489 |

Исходя из таблицы 3, наибольшей компрессии изображения удалось достигнуть с помощью формата .gif.

Сделаем вывод, что для хранения изображений с малой цветовой палитрой и передачи чёрно-белых изображений лучше всего подходит формат .gif. Для передачи цветного изображения по каналам электронных сетей подходит формат .jpeg.

Задание 2

Текст для сжатия: “Урааааааааааааа в атаку” (вариант 19). Текст содержит 23 символа, на каждый символ приходится 1 байт. Итого: 23 байта. Применим к сообщению кодирование методом RLE.

Коды символов в кодировке CP-1251:

* “У” – 21110 - 1101 00112
* “р” – 24010 – 1111 00002
* “а” – 22410 - 1110 00002
* “в” – 22610 - 1110 0010‬2
* “т” – 24210 - 1111 0010‬2
* “к” – 23410 - 1110 1010‬2
* “у” – 24310 – 1111 00112
* “ ” (пробел) – 3210 - 0010 00002

Упрощённое сообщение:

(1; “У”), (1; “р”), (13; “а”), (1; “ ”), (1; “в”), (1; “ ”), (1; “а”), (1; “т”) , (1; “а”), (1; “к”), (1; “у”)

Закодированное сообщение (14 байт):

**0**0000010 11010011 11110000 **1**0001101 11100000 **0**0001000 00100000 11100010 00100000 11100010 11110010 11100010 11101010 11110011

Коэффициент сжатия (по формуле 1.2): 164%.

Контрольная сумма: 010011012

Задание 3

Построим код Шеннона-Фано (таблица 4).

Таблица 4

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P |  | p(j) |  |  |  |  | Кол-во симв. | p(j)\*log2(p(j)) | l |
| а | 15 | 0,652173913 | 1 |  |  |  | 1 | -0,402176974 | 0,652173913 |
| \_ | 2 | 0,086956522 | 0 | 1 | 1 |  | 3 | -0,306396692 | 0,260869565 |
| р | 1 | 0,043478261 | 0 | 1 | 0 | 1 | 4 | -0,196676607 | 0,173913043 |
| в | 1 | 0,043478261 | 0 | 1 | 0 | 0 | 4 | -0,196676607 | 0,173913043 |
| т | 1 | 0,043478261 | 0 | 0 | 1 | 1 | 4 | -0,196676607 | 0,173913043 |
| к | 1 | 0,043478261 | 0 | 0 | 1 | 0 | 4 | -0,196676607 | 0,173913043 |
| у | 1 | 0,043478261 | 0 | 0 | 0 | 1 | 4 | -0,196676607 | 0,173913043 |
| У | 1 | 0,043478261 | 0 | 0 | 0 | 0 | 4 | -0,196676607 | 0,173913043 |

Энтропия равняется 1,888633307, а среднее число символов на одну букву - 1,956521739.

Закодированное сообщение:

0000 0101 1 1 1 1 1 1 1 1 1 1 1 1 1 011 0100 011 1 0011 1 0010 0001

Сгруппируем сообщение по байтам:

00000 10111111 11111111 01101000 11100111 00100001

Округлив, получаем 6 байт. Коэффициент сжатия (по формуле 1.2): 383%.

Задание 4

Построим код Хаффмана. Для построения кода заполним таблицу (табл. 5) и построим дерево (рис. 1)

Таблица 5

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P | p(j) |  |  |  |  |  |  |  |
| а | 0,652174 | 0,652174 | 0,652174 | 0,652174 | 0,652174 | 0,652174 | 0,652174 | 1 |
| \_ | 0,086957 | 0,086957 | 0,086956 | 0,086957 | 0,173914 | 0,173914 | 0,347828 |  |
| р | 0,043478 | 0,086957 | 0,086957 | 0,086957 | 0,086957 | 0,173914 |  |  |
| в | 0,043478 | 0,043478 | 0,086957 | 0,086957 | 0,086957 |  |  |  |
| т | 0,043478 | 0,043478 | 0,043478 | 0,086957 |  |  |  |  |
| к | 0,043478 | 0,043478 | 0,043478 |  |  |  |  |  |
| у | 0,043478 | 0,043478 |  |  |  |  |  |  |
| У | 0,043478 |  |  |  |  |  |  |  |

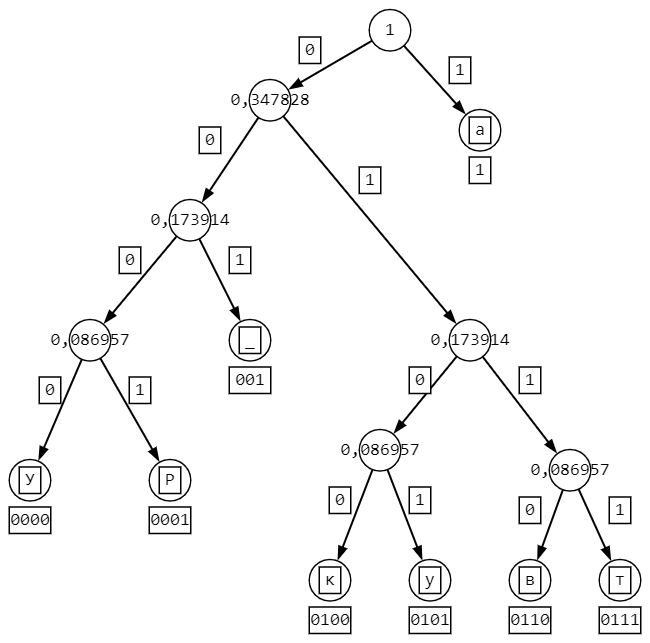


Рисунок 1. Дерево Хаффмана

Коды символов:

* “У” ­­– 0000
* “р” – 0001
* “а” – 1
* “в” – 0110
* “т” – 0111
* “к” – 0100
* “у” – 0101
* “ ” (пробел) – 001

Закодированное сообщение:

0000 0001 1 1 1 1 1 1 1 1 1 1 1 1 1 001 0110 001 1 0111 1 0100 0101

Сгруппируем сообщение по байтам:

00000 00111111 11111111 00101100 01101111 01000101

Количество бит такое же, как и при кодировании методом Шеннона-Фано. Округлив, получаем 6 байт. Коэффициент сжатия (по формуле 1.2): 383%.

Вывод

Методы сжатия данных, такие как RLE, Шеннона-Фано и Хаффмана, могут значительно уменьшить размер текстовых файлов. Проведенный анализ показывает, что для данного текста:

* RLE имеет наименьший коэффициент сжатия (164%), что указывает на относительно слабую эффективность сжатия. Метод эффективен для сжатия текста с большим количеством серий повторяющихся символов, в нашем случае серия всего одна.
* Методы Шеннона-Фано и Хаффмана показывают значительно более высокие коэффициенты сжатия (383%). Это означает, что они могут уменьшить размер файла примерно в 3,8 раза без потери информации за счёт использования однозначного кодирования.

В данном случае метод Шеннона-Фано и Хаффмана обеспечивают одинаковый уровень сжатия. Выбор между этими двумя методами обычно зависит от конкретных требований, таких как скорость сжатия и декомпрессии.