

## Задание № 1

### Построение частотных таблиц

#### *Описание задания*

Скопировать на свой рабочий диск архив с именем **Arh??.rar** (здесь и далее ?? — две цифры соответствующие номеру Вашего варианта). Разархивировать его. Получится 6 файлов: **Pic??.bmp**, **Pic??.jpg**, **Text??.doc**, **Text??.docx**, **Text??.txt** и сам архив **Arh??.rar**.

Создать программу (на языке Pascal или C) определяющую частоту появления каждого байта в файле.

Результаты исследования вывести в текстовые файлы с именем соответствующем расширению исследуемого файла + номер варианта и с расширением **tab**. Например, при обработке файла **Text01.doc** должен получиться файл с именем **doc01.tab**, а при обработке файла **Text01.txt** — файл **txt01.tab**.

#### *Содержание результирующего файла:*

1. первая строка — имя исследуемого файла;
2. вторая строка — количество байт в исследуемом файле;
3. пустая строка;
4. последующие строки — коды символов и частоты их появления, разделенные пробелом или, что желательнее, символом табуляции (код — 9), отсортированные по убыванию частот;
5. пустая строка
6. последняя строка — величина энтропии.

#### *Содержание отчета*

- 1) Тема
- 2) Описание задания
- 3) Текст программы для обработки и получения результата с комментариями, поясняющими основные моменты программы. Обязательно укажите среду, в которой была написана программа.

К отчету прилагаются результирующие файлы.

#### *Замечания*

1) Копируемые файлы имеют различные типы и размеры. Если на рабочем диске не хватает места для всех выбранных файлов, копируйте и обрабатывайте их по одному.

2) Так как байт может принимать 256 значений (от 0 до 255), то можно создать массив соответствующего размера. Это может быть массив записей с полями, хранящими код символа и количество его появлений в данном файле, или двумерный массив. Код каждого байта, вводимый в массив, вначале совпадает с его индексом.

3) Открывайте исследуемые файлы как файлы типа байт (BYTE). Считывая каждый очередной байт, добавляйте в соответствующий счетчик 1. Подсчитывайте и количество обработанных байтов.

4) После обработки всего файла необходимо данные отсортировать по убыванию показателей счетчиков любым известным вам способом.

5) Оценкой частоты появления каждого из байтов считайте отношение показания счетчика соответствующего этому байту к общему количеству байтов.

### *Пример*

После обработки файла с именем **Text01.txt**, получаем выходной файл с именем **txt01.tab**, который содержит следующие строки

Text01.txt  
20002

32 0.196530346965303  
238 0.0718928107189281  
232 0.0568943105689431  
224 0.0567943205679432  
229 0.0567443255674433  
237 0.0467453254674533  
235 0.0373462653734627  
...  
39 4.999500049995E-05  
40 4.999500049995E-05  
41 4.999500049995E-05  
57 4.999500049995E-05  
198 4.999500049995E-05  
219 4.999500049995E-05

4.68529550441711