

9. Файлы и сериализация

9.1. Файлы

9.1-0 В текстовом файле хранится таблица действительных чисел. В каждой строке записаны три числа, разделённые символом «;». Написать программу, находящую для каждого столбца сумму и арифметическое среднее.

9.1-1 Написать программу, оценивающую информационную энтропию в указанном файле.

Для этого она должна просмотреть файл побайтово и составить таблицу частот p_k каждого значения байта (k изменяется от 0 до 255). После этого можно вычислить энтропию по формуле

$$H = - \sum_{k=0}^{255} p_k \log_2 p_k.$$

Информационная энтропия в данном случае будет показывать количество информации, приходящейся на один байт файла. Чем она меньше, тем сильнее можно сжать файл.

В приведённой формуле в качестве p_k должны использоваться вероятности. Так как они, вообще говоря, неизвестны, вместо них используются частоты. Частота байта — это отношение количества его вхождений к общему количеству байтов в файле.

9.1-2 В текстовом файле записана информация о продолжительности телефонных звонков. Информация о каждом звонке записана в отдельной строке в формате «Фамилия:Продолжительность». Продолжительность указана в секундах.

Написать программу, находящую суммарную продолжительность разговоров для каждого из абонентов.

9.1-3 Написать функции для сохранения в двоичный файл и чтения из него двумерного массива целых чисел. В начало файла должны записываться размеры массива, а затем в двоичном виде элементы.

Элементы записываются построчно, то есть, вначале в файл помещаются все элементы первой строки массива, затем второй и так далее.

Функция для записи должна принимать имя файла и массив, функция для чтения — только имя файла и возвращать полученный массив.

Написать программу, использующую эти функции.

9.1-4 Написать программу, шифрующую файл и записывающую результат в новый файл. Шифрование выполнить прибавлением по модулю 2 («исключающее или», XOR) к каждому байту исходного файла заданного однобайтового числа.

Этот метод называется гаммированием. В случае наложения всего одного байта он крайне ненадёжен и не должен использоваться для обеспечения безопасности данных.

9.1-5 Написать программу, удваивающую каждый байт указанного двоичного файла с насыщением. Насыщение означает, что если удвоенное значение превышает 255 (максимальное значение для байта), то результат равен 255.

9.1-6 Написать программу, выводющую те строки текстового файла, в которых встречается указанное слово. Перед выводимой строкой указать её номер. В конце вывести общее количество вхождений слова.

9.1-7 Написать программу, просматривающую указанный файл побайтово и выводющую на экран 10 самых частых значений байтов с указанием частоты.

9.1-8 Написать программу, выводющую отсортированный список слов без повторений, встречающихся в указанном текстовом файле.

Считать словом последовательность букв, регистр символов не учитывать (для этого перевести все слова в нижний регистр). Остальные символы в файле игнорировать.

9.1-9 Написать программу, выводющую таблицу частот русских букв (без учёта регистра) в заданном текстовом файле. Таблицу упорядочить по убыванию частоты.

Частотой буквы считать отношение количества её вхождений к общему количеству букв.

9.2. Файловая система

9.2-0 Написать программу, дописывающую к имени всех файлов с расширением «*.txt» слева через пробел дату создания.

Дата должна быть записана в соответствии со стандартом ISO 8601: год, месяц и день через дефисы. Например: «2016-01-18» для 18 января 2016 года. Этот способ удобен тем, что позволяет сортировать даты естественным образом.

9.2-1 Написать программу, выводющую на экран список файлов с расширением «.txt» в заданной папке, в которых содержится указанное слово.

9.2-2 Написать программу, выводющую на экран все файлы в указанной папке с указанием размера. Также для каждого файла указать его долю в общем объёме.

9.2-3 Написать программу, которая объединяет содержимое всех файлов с именем, заканчивающимся на «.part.txt» и записывающая результат в файл с указанным именем. Файлы объединяются в алфавитном порядке.

9.2-4 Написать программу, выводющую список файлов, одновременно присутствующих в двух указанных папках. Сравнивать файлы по содержимому не требуется, достаточно искать файлы с одинаковыми именами.

9.2-5 Дан файл со списком имён файлов (по одному на строку). Написать программу, проверяющую, присутствуют ли файлы из списка в указанной папке. Имена отсутствующих файлов вывести на экран.

9.2-6 Написать программу, перемещающему каждый файл в папке в подпапку с именем, равным первому символу имени файла (без учёта регистра). Например, все файлы с именем, начинающимся с буквы «А» будут перемещены в папку «А» и т. д. (Папки должны создаваться программой при необходимости.)

9.2-7 Написать программу, выводящую на экран список всех файлов с расширением «.txt» в указанной папке. После каждого имени также должна выводиться первая строка файла.

9.2-8 Написать программу, выводящую список файлов в указанной папке, созданных раньше указанной даты.

9.2-9 Написать программу, перемещающую каждый файл в указанной папке в папку с именем, равным дате создания файла. Папки должны создаваться программой при необходимости.

Дата должна быть записана в соответствии со стандартом ISO 8601: год, месяц и день через дефисы. Например: «2016-01-18» для 18 января 2016 года. Этот способ удобен тем, что позволяет сортировать даты естественным образом.

9.3. XML и сериализация

9.3-0 Написать программу, запрашивающую список граждан (определяется фамилий, именем, отчеством и датой рождения) и записывающую его в файл в формате XML.

Написать программу, выводящую список из XML-файла.

9.3-1 Написать программу, получающую список файлов в указанной папке и сохраняющие их имена и даты создания в файл в формате XML.

Написать программу, выводящую список из XML-файла.

9.3-2 Написать программу, запрашивающую с клавиатуры фамилии и номера телефонов и сохраняющую их в файл в формате XML.

Написать программу для поиска номера телефона по фамилии в XML-файле.

9.3-3 Дан файл, хранящий переводы английских слов. Каждая пара перевод — слово занимает одну строку и записана в формате «слово - перевод».

Написать программу сохраняющую информацию о переводах в файл в формате XML.

9.3-4 Написать программу, запрашивающую с клавиатуры информацию о фильмах (название, режиссёр, год выхода) и сохраняющую её в файл в формате XML.

Написать программу, выводящую информацию о фильмах, сохранённую в XML-файл.

9.3-5 Дан файл в формате CSV. В первых трех столбцах записаны фамилия, имя и отчество, а в четвёртом — адрес электронной почты. Написать программу для преобразования файла в формат XML.

В формате CSV данные записываются построчно и разделяются запятыми. В случае, если запятая уже используется в качестве разделителя десятичной части числа (как в русском языке), используют точку с запятой.

9.3-6 Написать программу, запрашивающую информацию о книгах (автор, название и год издания) и сохраняющую её в файл в формате XML.

Написать программу, выводящую информацию о книгах, сохранённую в XML-файл.

9.3-7 Написать программу, запрашивающую информацию о группе (номер, год поступления и фамилии студентов) и сохраняющую её в файл в формате XML.

Написать программу, выводящую информацию о группе, сохранённую в XML-файл.

9.3-8 Дан файл, хранящий информацию о докладах на конференции. Информация о каждом докладе занимает одну строку и записана в формате «Фамилия: Тема (Дата)».

Написать программу для преобразования файла в формат XML.

9.3-9 Написать программу, запрашивающую информацию о студентах (фамилия и средний балл) и сохраняющую её в файл в формате XML.

Написать программу, принимающую на вход XML-файл и выводящую фамилии трёх лучших студентов.