

Задание практикума №2

Создание частотного словаря для некоторого текста

Написать программу, которая создает частотный словарь для заданного текста. Текст, по которому строится словарь, вводится со стандартного канала ввода, а построенный словарь выводится в стандартный канал вывода.

Входной текст нужно разбить на слова, в качестве разделителей считаем пробельные символы и концы строк. Для каждого слова нужно подсчитать число его вхождений в текст. Программа должна вывести в словарь все слова текста без повторений, упорядоченные по убыванию числа вхождений в текст. Формат выдачи:

слово1 *число-вхождений* *частота-вхождений-слова*
слово2 *число-вхождений* *частота-вхождений-слова*

.....

Например,

а	19	0.000432
в	12	0.000319
Москва	6	0.00016

.....

Частота вхождений слова — это отношение числа вхождений слова к общему числу слов в тексте.

Каждое слово выдается только один раз. Знаки препинания (точка, тире, запятая и т.д.), а также знаки типа вопроса, восклицания и т.п. считать отдельным словом. Слова считать различными, если есть хотя бы один различающийся символ. Приводить слова к одному регистру не требуется.

Про функции преобразования и классификации символов (<ctype.h>) можно прочитать в K&RC – приложение В, раздел 2.

Длина строк в программе и длина самого текста считаются неограниченными, но предполагается, что весь словарь должен поместиться в память. Также считаем, что входной текст обязательно заканчивается символом конца строки.

Для эффективности работы программы требуется, чтобы словарь хранился в бинарном дереве поиска, ключом служит само слово, информационной частью — счетчик вхождений.

Поскольку вводить тесты для программы вручную с клавиатуры затруднительно, равно как и анализировать вывод на экране, то рекомендуется использовать перенаправления ввода-вывода как для входного текста, так и для созданного словаря. Например, такая командная строка считывает входной текст из файла hotel.txt, а сгенерированный словарь записывает в файл hotel_dict.txt:

```
./task3 < hotel.txt > hotel_dict.txt
```