

Задание №3 для самостоятельного выполнения

Создание частотного словаря для некоторого текста

Написать программу, которая создает частотный словарь для заданного текста. Программа должна запускаться с помощью командной строки вида:

```
./prog -i f.txt -o dict.txt
```

Опция -i задает файл с входным текстом, опция -o задает выходной файл, в который должен быть выведен частотный словарь. При отсутствии опции -i ввод происходит из стандартного канала ввода, при отсутствии опции -o вывод происходит 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 прочесть в K&RC – раздел B2.

Длина строк в программе и длина самого текста считаются неограниченными, но предполагается, что весь словарь должен поместиться в память.

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