

Задание на знание языка C# (алгоритмическая задача)

Необходимо считать текст из файла и подсчитать частоту вхождения каждого слова.

В качестве слова рассматриваем непрерывную последовательность буквенных символов из латинского алфавита разделённую одним или несколькими пробелами. “Дефис” (минус -) считается буквой, остальные небуквенные символы рассматриваются как пробелы. Большие буквы отождествляются с маленькими (все буквы в последствии надо привести к маленьким).

Получившуюся последовательность пар вида «слово частота» необходимо сгруппировать по первой букве каждого слова и получившиеся группы отсортировать по алфавиту. Внутри каждой группы пары «слово частота» должны быть отсортированы по убыванию частоты.

Результат необходимо вывести в другой файл следующим образом: для каждой группы необходимо вывести:

1. Строка с буквой группы
2. Каждая пара внутри группы должна быть выведена на отдельной строке в виде: слово частота

Имя входного и выходного файла должны задаваться как параметры командной строки и могут быть произвольными. Примеры (для имени программы в примерах используется task.exe, у вас может быть что-то другое):

- task.exe input.txt result.txt
- task.exe i.txt o.txt

Пример входного файла:

Aaa aaa aa bbbb bbb bbb. Ddd ddd hh.

Для этого входного файла должен получиться следующий результат:

a

aaa 2

aa 1

b

bbb 2

bbbb 1

d

ddd 2

h

hh 1

Ожидается, что в результате работы над задачей, стажёр проявит навыки работы с файлами, строками и коллекциями в .Net. Использование LINQ и модульных тестов приветствуется и поощряется. Приёмная комиссия будет обращать внимание на именование переменных и организацию кода. При прочих равных условиях предпочтение будет отдано решению с более грамотной организацией кода.

Код решения задачи должен быть размещён в открытом репозитории на github и ссылка на этот репозиторий должна быть отправлена организаторам в качестве решения. При работе с репозитарием приёмной комиссией будет рассматриваться история комиттов.