

Тестовое задание для кандидата на должность разработчика программного обеспечения (Python)

Задача:

Написать функцию, которая определяет топ-10 самых популярных хэштегов в твитах и топ-5 самых частых слов для этих хэштегов. Хэштегом считается последовательность символов без пробелов, начинающаяся с символа '#'.

Входные данные:

Файл in.txt, кодировка utf-8

Содержит список строк, каждая из которых представляет собой последовательность символов (русские или английские буквы, цифры, знаки препинания), разделенные пробелами. Общая длина не более 140 символов.

Выходные данные:

Вывести в stdout два значения:

- Массив из 10 хэштегов, наиболее часто встречающихся во входных данных. Массив отсортирован по частоте в убывающем порядке.
- Словарь, где ключом является хэштег, значением - массив из 5 значимых слов, наиболее часто встречающихся в твитах с этим хэштегом. Значимым словом считается последовательность символов, не содержащая цифр или пробелов и начинающаяся с русских или английских букв. Словарь должен содержать списки слов только для топ-10 наиболее популярных хэштегов.

Язык разработки: python 2.7.x или python 3.6.x