Автоматизированный анализ объявлений о работе

Задача – анализ объявлений о работе (поиск ключевых навыков).

Представители различных профессий должны обладать различными навыками. Задача – определить список наиболее востребованных навыков смежных с основной профессией или технологией, указанной в поисковом запросе.

В качестве базы будем использовать hh.ru

Например, если задать на hh.ru запрос

<https://spb.hh.ru/search/vacancy?text=Программист&area=2&salary=&currency_code=RUR&experience=doesNotMatter&order_by=relevance&search_period=&items_on_page=500&no_magic=true>

то на 06.07.2017 получим 3 424 вакансии. Последовательно просматривая каждую вакансию получаем данные о ключевых навыках

Например:

|  |  |
| --- | --- |
| № | Навыки |
| 1 | Python, AngularJS, SQL, NoSQL, Git, Celery, Flask |
| 2 | Objective-C, iOS |
| 3 | 1С |
| 4 | JavaScript, jQuery, CSS3, Git, Java, HTML, HTML5, CSS, React, Redux |

Разбирая текст каждого из объявлений определяем – упоминался ли тот или иной навык в объявлении. Если навык упоминался, то ставим 1, если нет, то 0. Пример таблицы приведен ниже

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № объявл. | Навыки | | | | |
| Python | AngularJS | Java | SQL | … |
| 1 | 1 | 1 | 0 | 1 |  |
| 2 | 0 | 0 | 0 | 0 |  |
| 3 | 0 | 0 | 1 | 0 |  |

Сравнивая суммы по каждому из столбцов, определяем наиболее востребованные технологии.