

ТЗ на парсер

Задачи:

1. Спарсить результаты поисковой выдачи по 5 запросам.
2. Сложить их в таблицы.
3. Произвести расчеты показателей.

Сайт: LinkedIn.com

Раздел: Вакансии

Регион: США

Запросы:

1. Product Manager -growth*
2. Marketing Manager -growth*
3. Growth Manager -marketing, - product**
4. Growth Marketing -manager, -product**
5. Growth Product -manager, -marketing**

*Оба слова должны быть в названии вакансии, если в названии также содержится слово growth, то она не учитывается.

**Три запроса нужно собрать в одну таблицу. Принцип такой же: два слова в названии, не собирать, если есть минус слова. Смысл в том, чтобы охватить все возможные вакансии, содержащие слово growth.

В итоге должно получиться 3 таблицы без пересекающихся вакансий.

Фильтры по отраслям:

- Программное обеспечение
- Интернет-компании
- Интернет-медиа
- Информационные технологии и услуги
- Маркетинг и реклама

Искать нужно только по указанным выше отраслям.

Параметры таблицы:

1. Название вакансии
2. Название компании
3. Отрасль компании (одна из 5, добавленных в фильтр)
4. Количество сотрудников (1-10, 11-50, 51-200, 201-500, 501-1000, 1001-5000, 5001-10000, 10000+)
5. Должностной уровень (Молодой специалист, Специалист, Руководитель среднего звена, Директор, Другое)

6. Количество кандидатов на вакансию (1-10, 11-50, 51-200, 201-500, 500+)

Расчет по параметрам:

1. Количество вакансий в каждой таблице
 - 3 цифры
2. Количество вакансий по отраслям (для каждой таблицы)
 - 15 цифр
3. Количество вакансий в компаниях разного размера (для каждой таблицы)
 - 24 цифры
4. Количество вакансий по должностному уровню (для каждой таблицы)
 - 15 цифр
5. Количество вакансий по количеству кандидатов (для каждой таблицы)
 - 15 цифр

По итогам нужны:

- База данных с результатами парсинга
- Три CSV таблицы:
 1. Вакансии Product Manager
 2. Вакансии Marketing Manager
 3. Вакансии Growth
- Заполненная таблица с расчетами -
<https://docs.google.com/spreadsheets/d/1S0cxk14QApB0J8CS-kr-KF9QXeFiAkEoU7aSONa2tqA/edit?usp=sharing>

Расчеты будут перепроверены.