

# Построение путей

## 1 этап (парсинг данных и формирование первичных групп)

Данные парсим с HH по

Структура данных:

<https://docs.google.com/spreadsheets/d/1DEGS87CVkS1zCcvDAUetqEaZGgDrBFFJ6bQfPcqF0x8/edit?usp=sharing>

Ссылка на HH вакансии: <https://hh.ru/search/resume>

Id – выставляем для каждой строчки свой Id

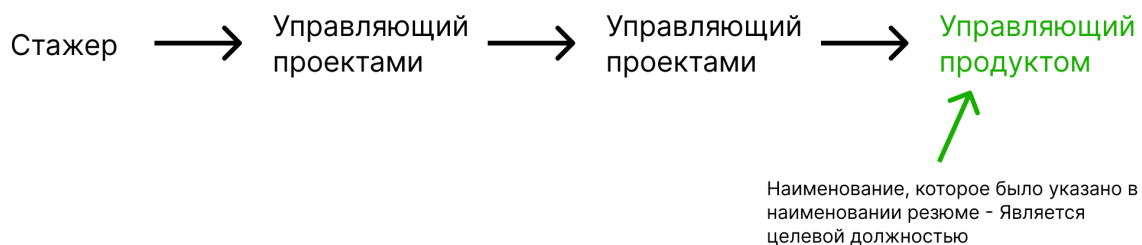
- № группы – один номер для строк, которые относятся к одному резюме
- Дата обновления – выставляем дату обновления вакансии. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Название профессии – забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Общий стаж - забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Специализация – забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- ЗП – забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Высшее образование (ВУЗ) – забираем из резюме из блока с образованием. Тут указываем то что выделено на сайте жирным шрифтом Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Высшее образование (направление) – забираем из резюме из блока с образованием. Тут указываем то что находится под жирным шрифтом. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Высшее образование (год окончания) – забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Знание языков – забираем из резюме. Первое тире в языке заменяем на скобку. Если указано несколько языков, то перечисляем через точку с запятой. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Опыт работы (промежуток) – забираем из резюме. Здесь указываем общий стаж как указано на сайте
- Месяцев – указываем сколько месяцев отработал кандидат по должности за которую отвечает данная строчка

- Должность – вставляем название должности из раздела с опытом работы, по которой работал кандидат
- Навыки – вставляем все навыки которые перечислены в резюме через точку с запятой. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Ссылка – указываем ссылку на резюме которое парсим. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Отрасль компании
- Подотрасль
- Специализация

**Все данные сохраняем в Яндекс Облако!**

## 2 этап (определение целевой профессии)

Целевой профессией, то есть профессия, которая будет находится в конце пути является должность, которая указана в заголовке резюме и располагается в спарсенной базе данных в столбце “Название профессии”



## 3 этап (определение повторяющихся групп)

Мы парсим данные раз в месяц. Есть вероятность, что некоторые из резюме будут повторятся. Поэтому для того чтобы не портить статистику такие группы нужно выявить и объединять

Группа считается идентичной если они удовлетворяют 5 условиям из 6 следующих:

1. Если группа имеет одинаковое название профессии (целевой профессии)
2. Если общий стаж двух резюме совпадает

3. Если группа имеет одинаковое кол-во этапов
4. Если должность указанная в определенном этапе одной группы равен должности того же этапа другой группы
5. Если длительность одного этапа одной группы равна длительность того же этапа другой группы
6. Если промежуток работы этапа в одной группе равен промежутку работы того же этапа в другой группе

При объединении двух групп:

1. Оставляем ту группу, которая была спарсена в более позднем промежутке парсинга
2. Оставляем ссылку на резюме, которая была спарсена в более позднем промежутке парсинга
3. Если у группы, которая была спарсена в более позднем промежутке парсинга появились новые данные о высшем образовании, знания языков, навыки, то мы добавляем их в оставшуюся группу

## **4 этап (определение уровней профессий в группе)**

Уровни профессий в группе определяем сравнивая названия в каждом этапе с базой данных: еще в разработке

<div> <div>Копия Соответствия и уровни</div> <div> <div>☆</div> <div>📁</div> <div>Сохранено на Диске.</div> </div> </div> <div> <div>Файл</div> <div>Правка</div> <div>Вид</div> <div>Вставка</div> <div>Формат</div> <div>Данные</div> <div>Инструменты</div> <div>Расширения</div> <div>Справка</div> <div>Последнее изменение: только...</div> </div> <div> <div> <div>↶</div> <div>↷</div> <div>🖨</div> <div>🔍</div> </div> <div>100%</div> <div>р. % .0 .00 123</div> <div>По умолча...</div> <div>10</div> <div> <div>B</div> <div>I</div> <div>U</div> <div>A</div> </div> <div> <div>🔗</div> <div>🔧</div> <div>🔍</div> <div>🔍</div> </div> <div> <div>☰</div> <div>⬇</div> <div>⬆</div> </div> </div> <div>F17</div> <div>fx</div> <div>Младший технический продакт-менеджер</div>						
	A	B	C	D	E	F
1	id	Соответствия	Вес профессии в соответствии	Уровень должности	Вес профессии в уровне	Наименование профессии и различные написания
2	1	1	1	0	0	Product manager
3	2	1	0	0	0	Управляющий продуктом
4	3	1	0	0	0	Продакт менеджер
5	4	1	0	0	0	Менеджер продукта
6	5	1	0	0	0	Менеджер по продукту
7	6	1	0	1	0	Ассистент Product-менеджера
8	7	1	0	1	0	Ассистент продуктового отдела
9	8	1	0	1	0	Помощник продукт менеджера
10	9	1	0	1	0	Ассистент менеджера по продукту
11	10	1	0	1	1	Junior Product Manager
12	11	1	0	1	0	Junior Product Marketing Manager
13	12	1	0	1	0	Junior Product Owner
14	13	1	0	1	0	младший менеджер продукта
15	14	1	0	1	0	младший продакт-менеджер
16	15	1	0	1	0	младший менеджер по продукту
17	16	1	0	1	0	Младший технический продакт-менеджер
18	17	1	0	2	1	Middle Product Manager
19	18	1	0	2	0	Менеджер продукта middle
20	19	1	0	3	1	Senior Product Manager
21	20	1	0	3	0	Старший продакт менеджер
22	21	1	0	3	0	Старший руководитель продуктов
23	22	1	0	3	0	Head of Product
24	23	1	0	3	0	Lead Product Manager
25	24	1	0	3	0	Head of Product
26	25	1	0	3	0	Senior Product Owner
27	26	2	0	0	0	Аналитик подготовки отчетов и данных
28	27	2	0	0	0	Аналитик данных
29	28	2	1	0	0	Data Analyst
30	29	2	0	0	0	Аналитик BI
31	30	2	0	0	0	Аналитик SQL
32	31	2	0	1	0	Младший аналитик данных
33	32	2	0	1	1	Junior Data Analyst
34	33	2	0	2	1	Middle Data Analyst
35	34	2	0	2	0	Middle Data
36	35	2	0	3	1	Senior Data Analyst

+

☰

Соответствия

Описания

Статьи

Книги

Youtube

Скидки

Для каждой профессии в каждом этапе группы выставляем определившийся уровень профессии, по следующей схеме:

- 0 - неопределена

- 1 - junior
- 2 - middle
- 3 - senior

Если уровень не был определен (возможно в том случае если такой профессии нет в базе данных) то помечаем профессию как неопределенную

Для целевой профессии определяем уровень так же по той же схеме.

## 5 этап (объединение повторяющихся этапов в группе)

Если в группе повторяются два, три или более этапов подряд, то их нужно объединить в один. При этом, стаж указанный в каждом этапе суммируется.

### В каких случаях и как производим объединение:

1. Название должности в одном этапе соответствует названию в следующем или предыдущем этапе (100% совпадение по написанию)

#### До объединения



#### После объединения





**ВАЖНО!** Если должности в двух этапах имеют 100% схожесть в названиях, но имеют разные отрасли и подотрасли компаний, то такие этапы не являются одинаковыми!

2. Если название должности в одном этапе, входит в одну группу и уровень должности в базе данных вариаций наименований профессий с наименованием должности в другой группе

<div> <div>Копия Соответствия и уровни</div> <div> <div>☆</div> <div>📁</div> <div>Сохранено на Диске.</div> </div> </div> <div> <div>Файл</div> <div>Правка</div> <div>Вид</div> <div>Вставка</div> <div>Формат</div> <div>Данные</div> <div>Инструменты</div> <div>Расширения</div> <div>Справка</div> <div>Последнее изменение: только...</div> </div>						
<div> <div>100%</div> <div>р. % .0 .00 123</div> <div>По умолча...</div> <div>10</div> <div> <div>B</div> <div>I</div> <div>U</div> <div>A</div> </div> <div> <div>🔍</div> <div>🔍</div> <div>🔍</div> <div>🔍</div> </div> <div> <div>☰</div> <div>☰</div> <div>☰</div> <div>☰</div> </div> <div> <div>⬇</div> <div>⬇</div> <div>⬇</div> <div>⬇</div> </div> <div> <div>⬆</div> <div>⬆</div> <div>⬆</div> <div>⬆</div> </div> </div>						
F17 <div>fx</div> Младший технический продакт-менеджер						
	A	B	C	D	E	F
1	id	Соответствия	Вес профессии в соответствии	Уровень должности	Вес профессии в уровне	Наименование профессии и различные написания
2	1	1	1	0	0	Product manager
3	2	1	0	0	0	Управляющий продуктом
4	3	1	0	0	0	Продакт менеджер
5	4	1	0	0	0	Менеджер продукта
6	5	1	0	0	0	Менеджер по продукту
7	6	1	0	1	0	Ассистент Product-менеджера
8	7	1	0	1	0	Ассистент продуктового отдела
9	8	1	0	1	0	Помощник продукт менеджера
10	9	1	0	1	0	Ассистент менеджера по продукту
11	10	1	0	1	1	Junior Product Manager
12	11	1	0	1	0	Junior Product Marketing Manager
13	12	1	0	1	0	Junior Product Owner
14	13	1	0	1	0	младший менеджер продукта
15	14	1	0	1	0	младший продакт-менеджер
16	15	1	0	1	0	младший менеджер по продукту
17	16	1	0	1	0	Младший технический продакт-менеджер
18	17	1	0	2	1	Middle Product Manager
19	18	1	0	2	0	Менеджер продукта middle
20	19	1	0	3	1	Senior Product Manager
21	20	1	0	3	0	Старший продакт менеджер
22	21	1	0	3	0	Старший руководитель продуктов
23	22	1	0	3	0	Head of Product
24	23	1	0	3	0	Lead Product Manager
25	24	1	0	3	0	Head of Product
26	25	1	0	3	0	Senior Product Owner
27	26	2	0	0	0	Аналитик подготовки отчетов и данных
28	27	2	0	0	0	Аналитик данных
29	28	2	1	0	0	Data Analyst
30	29	2	0	0	0	Аналитик BI
31	30	2	0	0	0	Аналитик SQL
32	31	2	0	1	0	Младший аналитик данных
33	32	2	0	1	1	Junior Data Analyst
34	33	2	0	2	1	Middle Data Analyst
35	34	2	0	2	0	Middle Data
36	35	2	0	3	1	Senior Data Analyst

+

☰

Соответствия

Описания

Статьи

Книги

Youtube

Скидки

Пример объединения:

#### До объединения



#### После объединения



В итоговом варианте выставляем наименование профессии из базы данных соответствий наименований профессий и уровней, наименования профессии из соответствующего уровня должности с весом профессии в уровне = 1



**ВАЖНО!** Если мы определили, что наименования профессий из двух разных этапов находятся в одной группе вариаций наименований профессий и в одном уровне профессий, но они имеют разные отрасли и подотрасли компаний, то такие профессии не являются одинаковыми и мы их не объединяем

База данных уровней профессий и вариаций наименований: еще в разработке

## 6 этап (определение похожих путей).

Пути считаются похожими если:

- Имеют одинаковый набор должностей в этапах в одинаковом порядке с учетом целевых должностей, но разными промежутками работы на данных должностях.



Стажер → Управляющий проектами → Управляющий продуктом → Управляющий продуктом

||

Стажер → Управляющий проектами → Управляющий продуктом → Управляющий продуктом

- Если у двух групп одинаковые этапы и эти этапы расположены в одинаковом порядке, но у одной из групп дополнительно прописаны дальнейшие этапы

- Если двух групп одинаковые этапы и эти этапы расположены в одинаковом порядке, но у одной из групп дополнительно прописаны

Этапы определения похожих путей:

1. По порядку выбираем отдельные группы и начинаем анализировать их этапы на похожесть с другими группами
2. Берем наименование должности 1-го этапа группы и сравниваем с другими первыми этапами

## 7 этап (определение развилок и альтернативных путей)

ё

Дмитрий Адамов приглашает вас на запланированную конференцию: Zoom.

Тема: Zoom meeting invitation - Zoom Meeting Дмитрий Адамов

Время: 1 апр. 2022 01:40 PM Москва

Подключиться к конференции Zoom

<https://us05web.zoom.us/j/87452914244?pwd=TTlVNlcyanpqTkRVUFdlWjdBc1grUT09>

Идентификатор конференции: 874 5291 4244

Код доступа: j0yUqq

в повышении квалификации нужно анализировать название на наличие слова “Стажировка”