

# 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

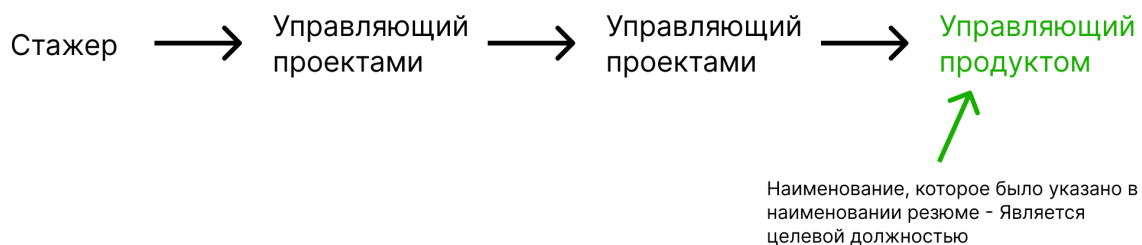
- № группы – один номер для строк, которые относятся к одному резюме
- Дата обновления – выставляем дату обновления вакансии. Эта ячейка дублируется для всех строк, которые относятся к одному резюме
- Название профессии – забираем из заголовка резюме. Эта ячейка дублируется для всех строк, которые относятся 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 нужной профессии, уровня и приведение к единому знаменателю)**

### **Определение нужной профессии**

В одном пути могут встречаться похожие друг на друга наименования у которых будет различия только в одну букву. Например: Project manager и Product manager. Так как это две совершенно разные профессии для нас важно чтобы система понимала это чтобы в дальнейшем мы могли привести их к одному знаменателю.

Для этого необходимо обратиться к таблице “Уровни профессий и вариации написания” и сравнить все наименования профессий в этапах резюме и целевой профессии с указанными в нулевом уровне группы соответствия.

Соответствия и уровни ☆ Сохранено на Диске.

Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка Последнее изменение: только...

100% 123 По умолча... 10

	A	B	C	D	E	F	G
1		Соответствия	Вес профессии в соответствии	Уровень должности	Вес профессии в уровне	Наименование профессии и различные написания	Профобласть
2	1	1	1	0	0	Product manager	Маркетинг, реклама, PR
3		1	0	0	0	Product-manager	
4	2	1	0	0	0	Управляющий продуктом	Маркетинг, реклама, PR
5	3	1	0	0	0	Продукт менеджер	Маркетинг, реклама, PR
6	4	1	0	0	0	Менеджер продукта	Маркетинг, реклама, PR
7	5	1	0	0	0	Менеджер по продукту	Маркетинг, реклама, PR
8	6	1	0	1	0	Ассистент Product-менеджера	Маркетинг, реклама, PR
9	7	1	0	1	0	Ассистент продуктового отдела	Маркетинг, реклама, PR
10	8	1	0	1	0	Помощник продукт менеджера	Маркетинг, реклама, PR
11	9	1	0	1	0	Ассистент менеджера по продукту	Маркетинг, реклама, PR
12	10	1	0	1	1	Junior Product Manager	Маркетинг, реклама, PR
13	11	1	0	1	0	Junior Product Marketing Manager	Маркетинг, реклама, PR
14	12	1	0	1	0	Junior Product Owner	Маркетинг, реклама, PR
15	13	1	0	1	0	младший менеджер продукта	Маркетинг, реклама, PR
16	14	1	0	1	0	младший продакт-менеджер	Маркетинг, реклама, PR
17	15	1	0	1	0	младший менеджер по продукту	Маркетинг, реклама, PR
18	16	1	0	1	0	Младший технический продакт-менеджер	Маркетинг, реклама, PR
19	17	1	0	2	1	Middle Product Manager	Маркетинг, реклама, PR
20	18	1	0	2	0	Менеджер продукта middle	Маркетинг, реклама, PR
21	19	1	0	3	1	Senior Product Manager	Маркетинг, реклама, PR

+ Соответствия Описания Статьи Книги Youtube Скидки



**ВАЖНО!** Учитывать при анализе возможно использование разных спряжений, родом в которых могут быть написаны профессии

При этом нужно понимать, что в этапах могут присутствовать наименования профессий из разных профобластей и из разных отраслей. Пример:

Стажер → Junior Project manager → Junior Product manager → UX исследователь → Junior Product manager

Следовательно каждое резюме необходимо анализировать по всем существующим в базе данных таблицам уровней профессий и вариаций написания. То есть, если после одного проведенного анализа наименований в этапах остались неопределенные наименования, то необходимо обращаться к следующей таблице и проводить такой же анализ до тех пор пока в пути не будут определены все наименования.



После того как мы определили профессии необходимо привязать их к одной из таблиц соответствий, уровней и вариаций написания для того чтобы в дальнейшем не приходилось проводить повторный анализ

## **Определение уровня профессии**

После того как были определены профессии необходимо определить их уровень. Для этого используем таблицу соответствий, уровней и вариаций написания к которым на предыдущем этапе мы привязали определенные наименования.

Копия Соответствия и уровни						
Сохранено на Диске.						
Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка Последнее изменение: только						
100% р. % .0 .00 123 По умолча... 10 B I A						
F17 fx 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

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.

### **Приведение названий профессий к одному знаменателю**

В названиях профессий как в этапах, так и в целевых профессиях могут встречаться разные вариации написания одной и той же профессии и ее уровня. Для того чтобы в дальнейшем нам было удобнее с ними работать необходимо приводить их к единому знаменателю.

Для этого обращаемся к таблице соответствий, уровне и вариаций написания и заменяем название которое указано в резюме на название, которое имеет 1 в столбце “Вес профессии в уровне”.

Пример: В резюме указано название “Ассистент продуктового отдела” меняем его на Junior Product Manager, которое имеет 1 в столбце “Вес профессии в уровне”.

Соответствия и уровни

Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка Последнее изменение: только...

100% р. % .0 .00 123 По умолча... 10 B I A

F9 fx Ассистент продуктового отдела

	A	B	C	D	E	F	G
		Соответствия	Вес профессии в соответствии	Уровень должности	Вес профессии в уровне	Наименование профессии и различные написания	Профобласть
1							
2	1	1	1	0	0	0 Product manager	Маркетинг, реклама, PR
3		1		0	0	0 Product-manager	
4	2	1		0	0	0 Управляющий продуктом	Маркетинг, реклама, PR
5	3	1		0	0	0 Продакт менеджер	Маркетинг, реклама, PR
6	4	1		0	0	0 Менеджер продукта	Маркетинг, реклама, PR
7	5	1		0	0	0 Менеджер по продукту	Маркетинг, реклама, PR
8	6	1		0	1	0 Ассистент Product-менеджера	Маркетинг, реклама, PR
9	7	1		0	1	0 Ассистент продуктового отдела	Маркетинг, реклама, PR
10	8	1		0	1	0 Помощник продукт менеджера	Маркетинг, реклама, PR
11	9	1		0	1	0 Ассистент менеджера по продукту	Маркетинг, реклама, PR
12	10	1		0	1	1 Junior Product Manager	Маркетинг, реклама, PR
13	11	1		0	1	0 Junior Product Marketing Manager	Маркетинг, реклама, PR
14	12	1		0	1	0 Junior Product Owner	Маркетинг, реклама, PR
15	13	1		0	1	0 младший менеджер продукта	Маркетинг, реклама, PR
16	14	1		0	1	0 младший продакт-менеджер	Маркетинг, реклама, PR
17	15	1		0	1	0 младший менеджер по продукту	Маркетинг, реклама, PR
18	16	1		0	1	0 Младший технический продакт-менеджер	Маркетинг, реклама, PR
19	17	1		0	2	1 Middle Product Manager	Маркетинг, реклама, PR
20	18	1		0	2	0 Менеджер продукта middle	Маркетинг, реклама, PR
21	19	1		0	3	1 Senior Product Manager	Маркетинг, реклама, PR

+ - Соответствия - Описания - Статьи - Книги - Youtube - Скидки



Нулевой уровень не трогаем!



При изменении наименований вся структура остается прежней. То есть мы оставляем прежними очередность этапов, сроки занимаемой должности в этапе, отрасли и подотрасли и все остальные данные относящиеся к данной группе

## 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:

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!



**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



Объединяем только похожие наименования в этапах. Не трогаем целевые профессии

## 6 этап (распределение профессий нулевого уровня по 1,2,3 уровню)

Распределение профессий нулевого уровня будет строиться на основе среднего значения профессионального опыта предшествующего занимаемой должности. То есть для нас важно понимать сколько пришлось отработать человеку до того как он занял одну из указанных в этапах должностей.

## Определение предшествующего опыта

Анализируем все имеющиеся в базе данных резюме, а именно этапы указанные в них и целевые профессии. Ищем первое упоминание профессии конкретного уровня в каждом из резюме.

### Пример:

Нам нужно определить средний профессиональный опыт с которым кандидат может претендовать на должность Junior Product Manager. Для этого мы обращаемся ко всей базе спарсенных резюме и ищем первое упоминание данной профессии.



Мы обращаемся именно ко всей базе спарсенных резюме, а не отдельной группе связанной именно с данной профессией



Может быть так, что первое наименование будет встречаться в целевой профессии. В таком случае мы берем общее значение всего опыта кандидата



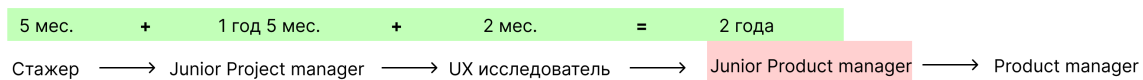
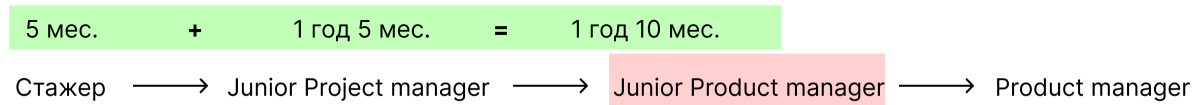
При анализе необходимо учитывать разные отрасли!

### Примеры возможных вариантов:

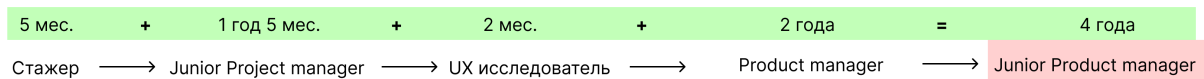
Когда путь в резюме начинается с искомой профессии. В таком случае считаем, что у кандидата вообще не было опыта (то есть 0)

Junior Project manager → Junior Product manager → UX исследователь → Junior Product manager

Когда искомая профессия находится на втором месте или на любом другом месте в пути резюме. В таком случае, сумируем весь предшествующей ей опыт.



Когда искомая профессия находится в целевой профессии. Берем весь опыт который есть у кандидата.



Если в резюме не найдено наименование из нужного уровня, то мы его пропускаем

## Вычисление среднего значения профессионального опыта для уровня

Сумируем все полученные значения в отдельности для каждого уровня и делим на количество данных значений.

## Определение предшествующего опыта каждой из профессий нулевого уровня.

Для того чтобы перераспределить профессии нулевого уровня по 1,2 и 3 уровням необходимо вначале вычислить предшествующий опыт относящийся к ним. Данный процесс строится по тому же принципу, как когда мы определяли предшествующий опыт для профессий разных уровней.

## **Перераспределение профессий нулевого уровня по 1,2 и 3 уровням**

Перераспределение строится согласно следующим условиям:

1. Если предшествующий опыт нулевой профессии входит в промежуток от нуля до среднего значения группы 1-го уровня, то данной профессии присваивается 1-й уровень и меняется название на название из таблицы уровней и вариаций написания с значение 1 в столбце “Вес профессии в уровне”
2. Если предшествующий опыт нулевой профессии входит в промежуток от среднего значения проф. опыта группы 1-го уровня до среднего значения группы 2-го уровня, то данной профессии присваивается 2-й уровень и меняется название на название из таблицы уровней и вариаций написания с значение 1 в столбце “Вес профессии в уровне”
3. Если предшествующий опыт нулевой профессии входит в промежуток от среднего значения проф. опыта группы 2-го уровня и выше то данной профессии присваивается 3-й уровень и меняется название на название из таблицы уровней и вариаций написания с значение 1 в столбце “Вес профессии в уровне”

## **7 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

Похожие пути объединяем в одну группу. Для этого в базе данных добавляем отдельный столбец "Соответствие пути" в котором для похожих путей указываем одинаковые значения. Пример: