

CV parse

трое в .KE

20.09.2020

Постановка задачи

- ▶ 2 таблицы
 - ▶ резюме
 - ▶ вакансии
- ▶ 3 модели
 - ▶ официант
 - ▶ кладовщик
 - ▶ водитель погрузчика
- ▶ выделить формализованную модель

Подход к решению

- ▶ Нет размеченных данных
 - ▶ сложно применять методы глубокого обучения и сложно интерпретировать результаты
- ▶ Данные большие и неструктурированные
 - ▶ разбить вопросы на общие и специфические
 - ▶ в ответах на общие вопросы содержатся ответы на специфические
 - ▶ данные не теряются при разбиении

Предложенная модель

Кладовщик (дано)		
Переменная	Описание	Тип данных
q1	Желаемая должность	string
q2	У вас есть опыт работы кладовщиком?	boolean
q2d	У вас есть опыт работы кладовщиком?	string
q3	Укажите Ваш стаж работы кладовщиком	integer
q4	Отрасль	string (array)
q5	Названия компаний	string (array)
q6	С какой системой хранения Вы имели опыт работы?	string (array)
q7	Опыт работы с программами	string (array)
q8	Опыт с инструментарием	string (array)
q9	Типы работ	string (array)
q10	Сколько времени вы потратили на поиск работы кладовщика?	integer
q11	Уровень ЗП (желаемый)	integer

Наша модель

Общие вопросы	
Id	Вопрос
q0	Название вакансии
q1	Регион
q2	Город
q3	Название компании
q4	Зарплата от
q5	Тип занятости
q6	Тип графика
q7	Опыт
q8	Время поиска работы (дни)
q9	Требования (образование, качества)
q10	Обязанности (что будет делать)
q11	Условия (условия работы)

Кладовщик	
Id	Вопрос
q0	Отрасль
q1	С какой системой хранения Вы имели опыт работы?
q2	Опыт работы с программами
q3	Опыт с инструментарием
q4	Типы работ

Реализация

- ▶ Наибольшая сложность – вольное изложение
- ▶ Разбитие описание вакансии
 - ▶ требования
 - ▶ обязанности
 - ▶ условия
- ▶ Каждую категорию разбить на смысловые единицы
- ▶ Каждая смысловая единица является потенциальным ответом на вопрос
 - ▶ не нужно искать ответы на вопросы во всём тексте
- ▶ Ставим задачу сопоставить вопросу смысловые единицы

Пример разбиения "description"(было)

- ▶ Обязанности: Прием товара (Кока-Кола) на склад от поставщиков; Распределение принятой продукции по адресам стеллажного хранения до 6 яруса (11,5 метров); Инвентаризация, внесение корректировок в базе данных; Прием возвратной продукции от водителей. Требования: Опыт работы кладовщиком от года; Опыт работы на электроштабеллере (желательно); Опытный пользователь ПК; Ответственность, внимательность. Условия: Оформление по ТК РФ; Своевременные выплаты заработной платы; График работы 1/1 с 20.00 до 20.00 (выходные суббота и воскресенье); Место работы: Московское шоссе 177 А (служебная развозка от м. Купчино, м. Пр.Ветеранов).

Пример разбиения "description"(стало)

- ▶ Обязанности
 - ▶ Прием товара (Кока-Кола) на склад от поставщиков
 - ▶ Распределение принятой продукции по адресам стеллажного хранения до 6 яруса (11,5 метров)
 - ▶ Инвентаризация, внесение корректировок в базе данных
 - ▶ Прием возвратной продукции от водителей
- ▶ Требования
 - ▶ Опыт работы кладовщиком от года
 - ▶ Опыт работы на электроштабеллере (желательно)
 - ▶ Опытный пользователь ПК
 - ▶ Ответственность, внимательность
- ▶ Условия
 - ▶ Оформление по ТК РФ
 - ▶ Своевременные выплаты заработной платы
 - ▶ График работы 1/1 с 20.00 до 20.00 (выходные суббота и воскресенье)
 - ▶ Место работы: Московское шоссе 177 А (служебная развозка от ст м Купчино, ст м Пр.Ветеранов)

Текущая реализация

- ▶ Для каждого вопроса существует облако тегов, соответствующих ответам на вопрос
- ▶ Ищем в смысловых единицах вхождения тегов
- ▶ В качестве ответа на вопрос возвращаем смысловую единицу с соответствующим вхождением
- ▶ Таким образом заполняем ответы на специфичные вопросы

Пример тегов для кладовщика

☐ A

B

C

1	Кладовщик		
2	Id	Вопрос	Пример тега
3	q0	Отрасль	"производство", "завод"
4	q1	С какой системой хранения Вы имели опыт работы?	"Адресная", "Стеллажная"
5	q2	Опыт работы с программами	"1С", "Microsoft", "ПК",
6	q3	Опыт с инструментарием	"штабелер", "рохля", "складской техники"
7	q4	Типы работ	это и есть "обязанности" из общих вопросов

Результат работы алгоритма

```
"17746005": {
  "common": {
    "q0": "Кладовщик",
    "q1": "Санкт-Петербург",
    "q2": "посёлок Шушары",
    "q3": "4Ф",
    "q4": "38000.0",
    "q5": "Полная занятость",
    "q6": "Сменный график",
    "q7": "От 1 года до 3 лет",
    "q8": "1499",
    "q9": [
      "Опыт работы кладовщиком от года",
      "Опыт работы на электроштабелере (желательно)",
      "Опытный пользователь ПК",
      "Ответственность, внимательность."
    ],
    "q10": [
      "Прием товара (Кока-Кола) на склад от поставщиков",
      "Распределение принятой продукции по адресам стеллажного хранения до 6 яруса (11,5 метров)",
      "Инвентаризация, внесение корректировок в базе данных",
      "Прием возвратной продукции от водителей."
    ],
    "q11": [
      "Оформление по ТК РФ",
      "Своевременные выплаты заработной платы",
      "График работы 1/1 с 20.00 до 20.00 (выходные суббота и воскресенье)",
      "Место работы: Московское шоссе 177 А (служебная развозка от ст м Купчинно, ст м Пр.Ветеранов)."
    ]
  },
  "special": {
    "q0": [],
    "q1": [
      "Распределение принятой продукции по адресам стеллажного хранения до 6 яруса (11,5 метров)"
    ],
    "q2": [
      "Опытный пользователь ПК"
    ],
    "q3": [],
    "q4": [
      "Прием товара (Кока-Кола) на склад от поставщиков",
      "Распределение принятой продукции по адресам стеллажного хранения до 6 яруса (11,5 метров)",
      "Инвентаризация, внесение корректировок в базе данных",
      "Прием возвратной продукции от водителей."
    ]
  }
},
}
```

Преимущества

- ▶ Решение полностью работает для таблицы вакансий
- ▶ Оно масштабируемое
- ▶ Быстро работает
- ▶ Не требовательно к вычислительным ресурсам
- ▶ Теги уже есть на многих сайтах по поиску работы

Что можно добавить

- ▶ Искать включения без тегов по расстоянию между вопросом и смысловой единицей
- ▶ Выводить вопросы, на которые не были получены ответы

Контакты

- ▶ @taya_penskaya
- ▶ @furiousteabag
- ▶ @Libkneht