

Автоматизация процесса подбора персонала с помощью математического моделирования.

**Исследовательская ВКР,
Копчев В.**

2022

Тема, объект и предмет исследования

- Исследовательская ВКР
- Тема: Автоматизация процесса подбора персонала с помощью математического моделирования.
- Объект: задача автоматизация отбора персонала с помощью ML
- Предмет: алгоритмы ML, позволяющие анализировать резюме (нейросети, кластеризация, текстовая аналитика)
- Цель: изучить алгоритмы, разработать свою систему отбора персонала

Актуальность темы

- Вакансии аналитиков данных становятся все более популярными, мы видим постоянные рекламы курсов
- Необходимо справляться с большим количеством соискателей
- Можно применить методы NLP и ML для автоматизации отбора персонала
- 500 вакансий (<https://career.habr.com/vacancies/analitik?ysclid=lbdi4jt8sd440109047>), 30000 резюме (собрано мной)

План работы

- Собрать данные с hh.ru
- Извлечь из них полезные сущности
- Произвести текстовую аналитику: кластеризация, популярные слова в опыте работы у соискателей, образование
- Отсекать соискателей из «плохих» кластеров

Релокация?	Командировка?	Город?	Аналитические комбинации слов	Справочник по интересам
prepared for occasional business trips	willing to relocate	Moscow	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	---
готова к редким командировкам	не готова к переезду	Москва	---	found smt by rule (Analyst) (Аналитик данн
готова к командировкам	не готова к переезду	Москва	---	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	готов к переезду	Санкт	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
не готов к командировкам	не готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	---
готов к командировкам	готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам		Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
	not willing to relocate	Moscow	---	found smt by rule (Analyst) (Аналитик данн
	willing to relocate	Moscow	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
не готов к командировкам	не готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	готов к переезду	Пенза	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	не готов к переезду	Москва	---	found smt by rule (Analyst) (Аналитик данн
готов к редким командировкам	не готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готова к командировкам	готова к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	готов к переезду	Санкт	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к редким командировкам	не готов к переезду	Москва	---	found smt by rule (Analyst) (Аналитик данн
готова к редким командировкам	готова к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к редким командировкам	не готов к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готова к редким командировкам	не готова к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
не готова к командировкам	не готова к переезду	Москва	found smt by rule (SQL) (Excel) (Python) (Tableau) (Power BI)	found smt by rule (Analyst) (Аналитик данн
готов к командировкам	не готов к переезду	Москва	---	found smt by rule (Analyst) (Аналитик данн

Обзор литературы

- I. Gusev “Dataset for Automatic Summarization of Russian News”. Исследование автоматического вывода краткого содержания текста на русском языке на примере датасета Gazeta, сравнение моделей, доказательство того, что mBART хорошо справляется с задачей.
- Y. Liu and others “Multilingual Denoising Pre-training for Neural Machine Translation”. Статья, описывающая модель mBART, ее работу на различных датасетах и для различных задач машинного перевода. Модель seq2seq для очищения текста от шума. Лучше подходит для некоторых специфических задач MT (в отличие от BART) и предобучена на текстах на нескольких языках
- M. Lewis and others “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”. Нейросеть с архитектурой Transformer BART — seq2seq, очищение текста от шума.