

# Большие данные для информационной системы рынка труда (ИСРТ) Формирование системы анализа спроса с использованием данных об онлайн-вакансиях

Вебинар, 2 июня 2020 г

Предварительные результаты анализа спроса на умения  
Другие варианты консолидированного использования данных и примеры

Сессия 2

Формирование системы больших данных для ИСРТ

Алессандро Ваккарينو (Alessandro Vaccarino)



# КАК ЭТО СДЕЛАТЬ?



# ПОЛУЧЕНИЕ ДАННЫХ — ПРАВИЛЬНЫЙ ВЫБОР ИСТОЧНИКОВ

Ряд источников доступен на рынке труда онлайн:

- Государственные/международные порталы
- Отраслевые/общие
- Веб-сайты компаний/газеты/агрегаторы/государственные службы занятости/...
- Социальные сети
- ...



Как выбрать наиболее подходящий и как их ранжировать?

# ПОЛУЧЕНИЕ ДАННЫХ — ОПРЕДЕЛЕНИЕ И РАНЖИРОВАНИЕ РЕСУРСОВ

С поддержкой экспертов рынка труда (РТ) в конкретных странах, по которым был проведен анализ, был составлен список наиболее актуальных порталов с целью обеспечения надежности и качества проанализированных источников.

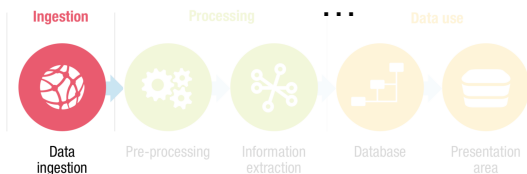
Количество опубликованных ОВ

Тип портала

Структурированные/неструктурированные данные

Рейтинг веб-страниц в Google (проверенное использование)

| Final Ranking |  |
|---------------|--|
| 2             |  |
| 4             |  |
| 8             |  |
| 7             |  |
| 6             |  |
| 12            |  |
| 13            |  |
| 11            |  |
| 1             |  |
| 3             |  |
| 10            |  |
| 5             |  |



# ПОЛУЧЕНИЕ ДАННЫХ — СБОР ДАННЫХ — ЦЕЛЬ...

**JUNIOR SOFTWARE DEVELOPER**

Location: United Kingdom  
Application deadline: Saturday, 30 September 2017  
Reference number: 100

Home > Now Hiring: Software Developers > Junior Software Developer

Share: [in](#) [t](#) [g+](#) [e](#)

**Description**

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful APIs, and third party integrations.

We're looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

Название:

Младший разработчик программного обеспечения

Область:

Великобритания

Время:

Суббота, 30 сентября 2017 г

Описание:

Младший разработчик программного обеспечения отвечает за разработку программного обеспечения высокого качества для использования...



# ПОЛУЧЕНИЕ ДАННЫХ — СБОР ДАННЫХ — ...СТРАТЕГИЯ...



Акцент на  
объемах



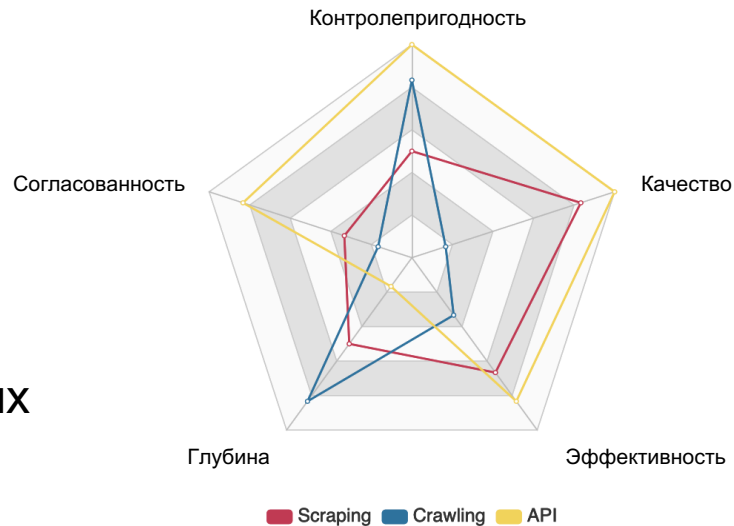
Приумножение и  
максимальное  
увеличение охвата



# ПОЛУЧЕНИЕ ДАННЫХ — СБОР ДАННЫХ — ...МЕТОДИКИ

Индивидуальный подход с целью максимального повышения эффективности:

- «деликатное» получение данных (без перегрузки порталов)
- Извлечение определенных признаков при их наличии (сектор, место и т. д.)
- Масштабирование при необходимости



Отслеживание Перемещение Сбор, обработка и интерпретация данных



# ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ — ЦЕЛЬ...

После завершения этапа получения данных осуществляется сбор большого количества веб-страниц. Но:

- Сколько из них действительно являются ОВ?
  - Нежелательная информация (профили в социальных сетях, страницы с политиками и т. д.)
  - «Фейковые» вакансии (тренинги, консалтинговые услуги и т. д., публикуемые на порталах ОВ)
- Сколько из них являются уникальными ОВ?



# ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ — ...СТРАТЕГИЯ...



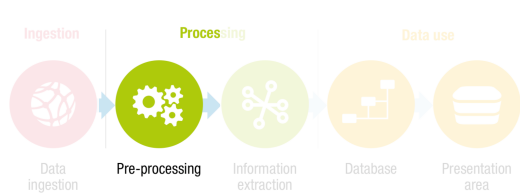
Определение  
стиля



Уменьшение  
количества  
нежелательной  
информации



ОВ  
Дубликаты



# ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ — ...МЕТОДИКИ

Для обнаружения нежелательной информации и дублирующих ОВ используется несколько методик с применением искусственного интеллекта (например, спам-фильтр).

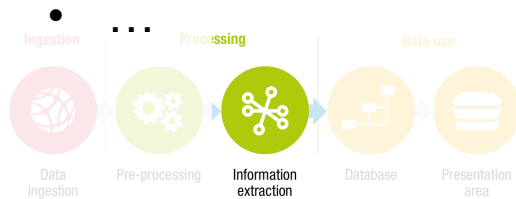
Определение стиля — первый этап, целью которого является акцент на определенных формулировках.



# КЛАССИФИКАЦИЯ ДАННЫХ — ЦЕЛЬ...

После завершения этапа предварительной обработки данных мы получаем релевантное количество онлайн-вакансий. Теперь самое время **извлечь информацию**:

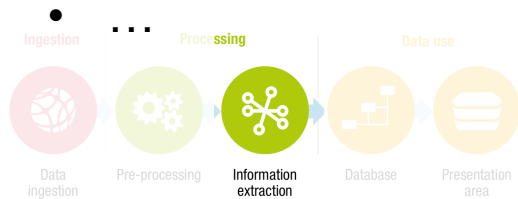
- Какие профессии нужны?
- Указывается ли место работы?
- Относится ли работа к определенному сектору?
- Какие умения требуются?



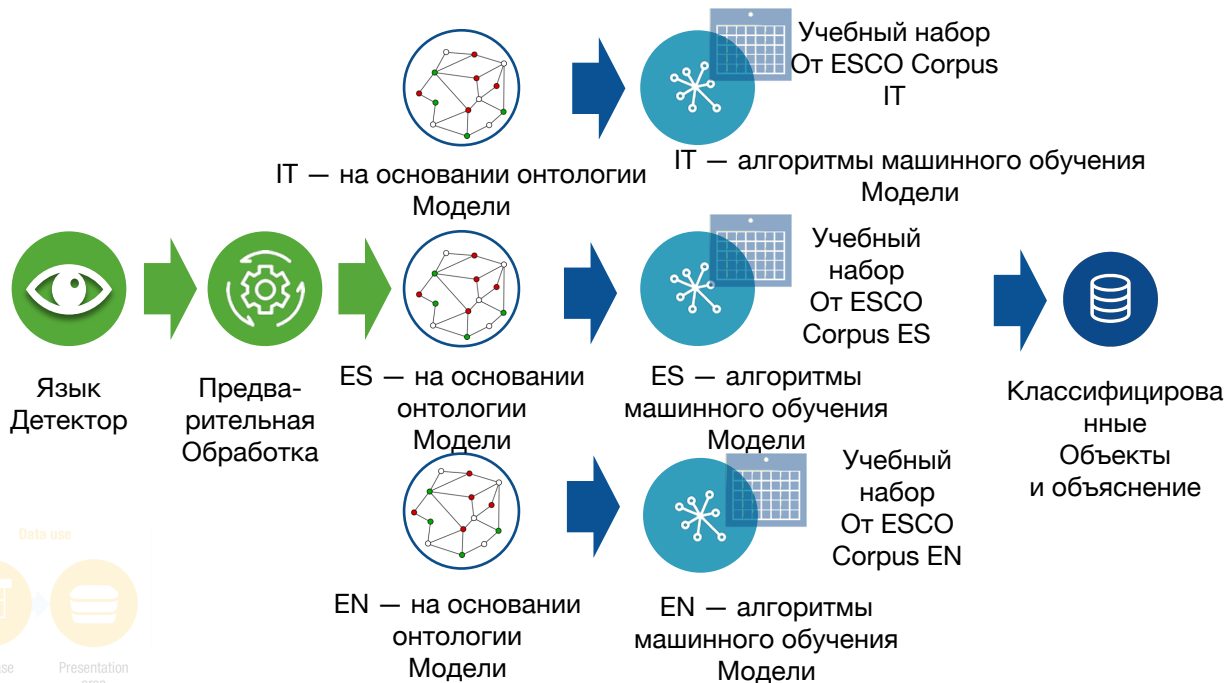
# КЛАССИФИКАЦИЯ ДАННЫХ — ...СТРАТЕГИЯ...

Классификация и систематизация для каждой аналитической переменной для установки области значений:

- Профессия (ESCO)
- Умения (ESCO)
- Место (NUTS/ISO3166)
- Сектор (NACE)
- Уровень образования (ISCED)



# КЛАССИФИКАЦИЯ ДАННЫХ — ...МЕТОДИКИ



# ИСПОЛЬЗОВАНИЕ И ПРЕДСТАВЛЕНИЕ ДАННЫХ

На **сессиях 3 и 6** будет представлено несколько примеров использования



# СПАСИБО!

