

СОЗДАНИЕ ИНФОРМАЦИОННОГО СЕРВИСА ДЛЯ МОНИТОРИНГА РЫНКА ТРУДА НА ОСНОВЕ ДАННЫХ HEADHUNTER

Е.С. Колесников

бакалавр

Г.С. Завалин

магистрант

*Владивостокский государственный университет
Владивосток, Россия*

Умение оперативно трансформировать образовательные программы под быстро меняющиеся требования рынка труда является одним из залогов успешного трудоустройства выпускников вузов. Целью данной работы является разработка информационного сервиса для мониторинга рынка труда в разрезе направлений подготовки высшего образования ВВГУ на основе данных популярной платформы для размещения вакансий HeadHunter.

Ключевые слова: рынок труда, мониторинг, дашборд, анализ данных, python, grafana

CREATION OF AN INFORMATION SERVICE TO MONITOR THE LABOR MARKET BASED ON HEADHUNTER DATA

The ability to promptly transform educational programs to meet the rapidly changing requirements of the labor market is one of the keys to the successful employment of university graduates. The purpose of this work is to develop an information service to monitor the labor market in the context of higher education areas of VVSU based on data from the popular platform for posting vacancies HeadHunter.

Keywords: market, monitoring, dashboard, data analysis, python, grafana

В современном мире, когда технологии развиваются с невероятной скоростью, сложно уследить за ними и знать, чему надо обучаться для успешного трудоустройства. Высшее же образование зачастую поддается критике из-за своей оторванности теоретических знаний от практики. Для актуализации учебных программ университету необходимо всегда опираться на рынок труда – именно он диктует, какие профессии нужны региону и какими навыками должны обладать сотрудники. Для решения этой проблемы и поддержания актуальных сведений по рынку труда был разработан информационный сервис для мониторинга вакансий с площадки HeadHunter. В основе сервиса лежит возможность фильтрации вакансий и навыков по направлениям подготовки Владивостокского государственного университета для определения популярных профессий и навыков, которые необходимы студентам, обучающимся каждый на своем направлении.

Платформой для сбора данных был выбран сайт HeadHunter (НН) – одна из крупнейших платформ по поиску работы и сотрудников, позволяющая работодателям быстро выбрать подходящих работников, а соискателям – искомую работу [1]. НН обладает открытым API (Application Programming Interface), с помощью которого можно обмениваться данными с сервисами HeadHunter (сбор вакансий, их размещение, просмотр резюме и т.д.). Интерфейс – это граница между двумя функциональными системами, на которой происходит их взаимодействие и обмен информацией. При этом процессы внутри каждой из систем скрыты друг от друга. С помощью интерфейса можно использовать возможности разных систем, не задумываясь о том, как они обрабатывают наши запросы и что у них «под капотом» [2].

Поиск вакансий посредством API работает аналогично поиску на сайте НН – вводим ключевое слово (навык или профессию) и соответствующий метод API отправляет запрос на сервер и возвращает набор вакансий согласно запросу. Для сбора вакансий использовались предварительная разметка профессий в разрезе направлений подготовки – бизнес-роли, которые получают студенты по итогу обучения.

Для сбора данных и их обработки использовался язык программирования Python, так как он имеет множество полезных библиотек. Основные из них это Pandas для работы с данными,

requests для запросов к API, pyodbc для запросов к БД через Python, pymorphy2 для работы с текстом. Для хранения данных собранных данных была создана база данных (БД) на основе системы управления базами данных MSSQL. Первым этапом идет сбор данных, для этого был разработан алгоритм (рис. 1): набор данных (вакансий) нужно для начала получить в формате JSON (JavaScript Object Notation), после чего извлечь данные в пригодный для анализа формат, определить профессию вакансии, выявить ключевые навыки и занести все собранные данные в БД. Важной частью данного алгоритма является определение профессии по заголовку вакансии с помощью сходства Джаро-Винклера – меры схожести строк для измерения расстояния между двумя последовательностями символов. Это вариант, который в 1999 году предложил Уильям Э. Винклер. Неформально, расстояние Джаро между двумя словами – это минимальное число одно – символьных преобразований, которое необходимо для того, чтобы изменить одно слово в другое [3].

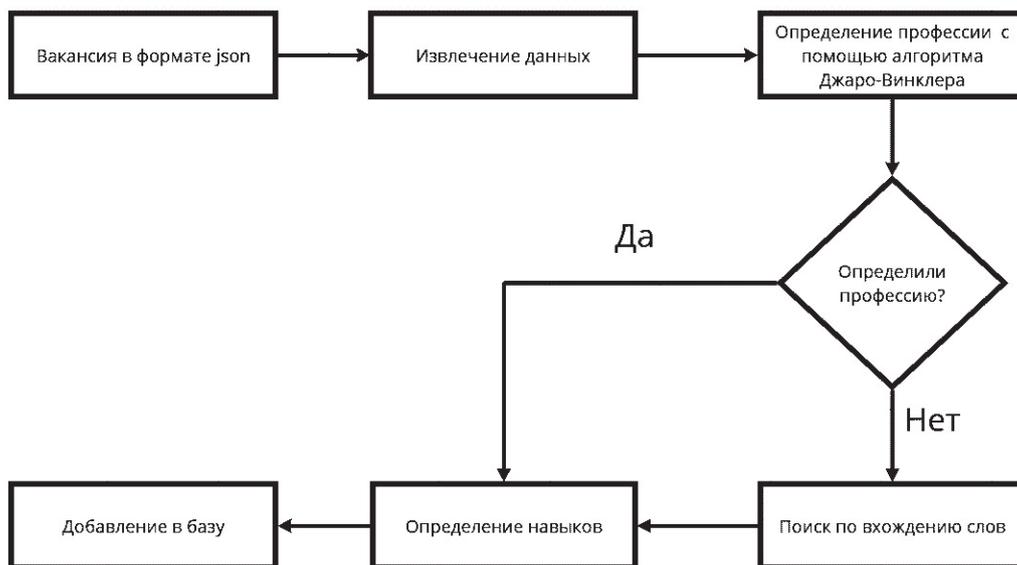


Рис. 1. Алгоритм обработки вакансий

В дальнейшем планируется построить классификатор для определения профессии, указанной в вакансии, так как алгоритм расчета с помощью расстояния Джаро иногда ошибочно привязывает не ту профессию к вакансии.

Навыки, которые работодатель требует от соискателей, находятся в тексте вакансии и требуют особой программной обработки. Для определения навыков из описания был разработан алгоритм на языке программирования Python (рисунок 2), который разбивает текст на слова (токенизация), лемматизирует (приводит к единой форме) их и сверяет с навыками из справочника. Справочник навыков содержит в себе уже заранее выявленные навыки и их разметку на hard, soft и digital skills.



Рис. 2. Алгоритм определения навыков

Название вакансии	Профессия	Описание вакансии	Навыки
Бухгалтер на банк-клиент (владивосток)	Бухгалтер	Обязанности: - Обработка заявок на оплату, проведение платежей по группе компаний, в том числе валютных - Взаимодействие с банками по вопросам предоставления валютных документов - Выгрузка банковских выписок в 1С, контроль корректности разнесения платежей (статьи ддс, заявки, договоры) - Ответы на запросы из банков, заказы справок, обновление ключей банковских сертификатов - Выгрузка банковских выписок с УТ в БГП, контроль корректности остатка по счету 51 в Б - Автоматизация бизнес процессов на своем участке - Выполнение поручений главного бухгалтера. Требования: - Профильное высшее образование - Опыт работы в аналогичной должности от 3 лет - Опыт работы с валютными платежами - Умение работать с оргтехникой - Знание офисных программ (MS Word, MS Excel), уверенный пользователь 1С 8.3, 8.2 - Внимательность, ответственность, стрессоустойчивость, умение работать с большим объемом информации, самодисциплина и организованность Работа с большим объемом информации Банк-клиент 1С: Бухгалтерия MS Outlook 1С: Бухгалтерия	MS Word 1С Высшее образование MS Excel Ддс Самодисциплина Автоматизация Валютные платежи Бухгалтерия Контроль Организованность Бизнес процессы Внимательность Стрессоустойчивость MS Outlook Ответственность
Ведущий менеджер по международным перевозкам, федеральная логистическая компания.	Менеджер по международным перевозкам		
Директор магазина	Директор		
Ведущий инженер / руководитель проекта (офис владивосток)	Руководитель проектов		
Менеджер по маркетингу и pr/event-маркетолог	Менеджер по маркетингу		
Руководитель отдела маркетинга, рекламы и связи с общественностью	Руководитель отдела маркетинга		
Нг-директор (логистика, таможня)	Нг директор		
Менеджер по работе с клиентами (удаленно)	Менеджер по работе с клиентами		
Event-менеджер по работе с партнерами	Менеджер по работе с партнерами		
Маркетолог в сферу жож и пп	Маркетолог		

Рис. 3. Пример обработки текста

Пример работы обоих алгоритмов представлен на рис. 3. В том случае, если не удается определить профессию по названию вакансии, то данное название вакансии будет добавлено в справочник профессий и будет в дальнейшем использоваться.

На данный момент собрано более 300 тысяч вакансий по различным профессиям, выявлено 81480 компаний. Собранные данные были представлены в виде информационного экрана (дашборда), который позволяет фильтровать собранные вакансии по профилю подготовки и региону (рис. 4).

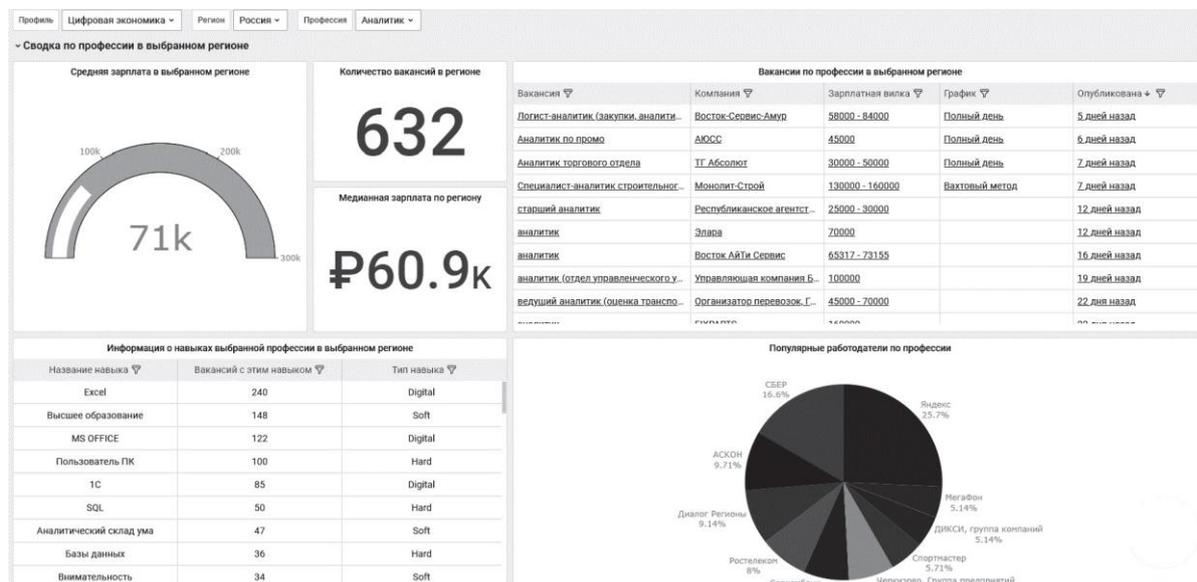


Рис. 4. Дашборд

На дашборде имеется возможность посмотреть статистику по определенной профессии: среднюю заработную плату, популярных работодателей и самое главное – набор популярных навыков, которые формируют эту профессию. Для руководителей образовательных программ и директоров институтов данный инструмент может быть полезен для актуализации учебных программ – включения новых дисциплин, которые будут формировать у студентов актуальные навыки, необходимые для успешного трудоустройства.

Таким образом, был разработан алгоритм сбора данных из вакансий с ведущей платформы HeadHunter, в рамках которого реализовано обработка текстовой информации: определение профессии по заголовку и выявление навыков из описания вакансии. Для хранения данных была организована структура базы данных на основе СУБД MSSQL. Для пользователей внутри университета был создан дашборд с визуализацией данных на базе платформы Grafana, которая позволяет предоставлять пользователям доступ к дашбордам согласно ролевой модели.

-
1. Описание системы HeadHunter. – Текст: электронный. – URL: <https://soware.ru/products/headhunter>
 2. API – что это такое. – Текст: электронный. – URL: https://skillbox.ru/media/code/chto_takoe_api/
 3. Сходство Джаро-Винклера. – Текст: электронный. – URL: <https://infostart.ru/1c/articles/1172479/>