

Ливинская В.А., Гайчуков Е.И.
Белорусско-Российский
университет
г. Могилёв, Республика Беларусь

ИСПОЛЬЗОВАНИЕ ИНТЕРНЕТА ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА РЫНКА ТРУДА В ИТ-СФЕРЕ

Аннотация. Исследование посвящено анализу вакансий специалистов IT-профиля, размещенных на сайте HeadHunter. Для сбора информации из веб-ресурсов использовались библиотеки языка Python и программная среда RSTUDIO для проведения достоверного статистического анализа. Для определения факторов, оказывающих влияние на величину начального уровня заработной платы, был проведен дисперсионный анализ Фишера.

Ключевые слова: рынок труда, дисперсионный анализ, разработка программного обеспечения.

Abstract. This study is about the analysis of vacancies of IT-profile specialists posted on the HeadHunter website. Information from the web resources was collected using the Python programming language. The RSTUDIO software environment was used to carry out reliable statistical analysis. Fisher's analysis of variance was carried out to determine the factors that influence the value of the initial wage level.

Keywords: labor market, analysis of variance, software development.

Одной из основных движущих силой роста экономики любой страны являются трудовые ресурсы: их количество, профессиональные навыки, возрастной состав, концентрация на определенной территории. Традиционно, этими вопросами занимается социально-экономическая статистика, инструментами которой являются абсолютные и относительные показатели, их динамика за определенный период.

Для проведения качественного анализа рынка труда, позволяющего определять достоверность различий, например, заработных плат у категорий работников, проживающих на различных территориях, имеющих одинаковую квалификацию, влияние определенных компетенций на величину заработной платы недостаточно данных, размещаемых с годичной или квартальной периодичностью в официальных статистических сборниках, в силу того, что данные, публикуемые в этих сборниках являются агрегированными, не отражающими нюансы реальной ситуации на рынке. Оптимальными данными можно считать данные, отражающие процессы, проходящие на рынке труда в реальном времени. Для анализа таких данных необходимо выбрать источник их сбора и инструменты для систематизации, хранения и обработки. Ключевыми параметрами для организации мониторинга рынка труда должны высту-

пать: совокупность спроса и предложений, время и место исследования. Поскольку любые выводы с применением методов прикладной статистики делаются на основании репрезентативных выборочных совокупностей, в качестве объекта наблюдений могут использоваться объявления о имеющихся вакансиях (модель спроса на рынке труда) и объявления соискателей в виде резюме (модель предложения).

Современные информационные технологии позволяют организовать мониторинг рынка вакансий почти в реальном времени, путем сбора информации с помощью парсинга сайтов-агрегаторов с объявлениями о размещаемых вакансиях.

Данное исследование посвящено анализу вакансий специалистов IT-профиля, размещенных на сайте HeadHunter, одном из самых крупных ресурсов по поиску работы и сотрудников в мире (по данным рейтинга Similarweb). По независимым оценкам, сайт имеет 49 млн резюме в базе, 793 тыс вакансий, 1423 тыс проверенных компаний посещают 18 млн посетителей ежемесячно, оставляя около 1 млрд откликов на вакансии. Исследования рынка труда проводятся такими гигантами как Superjob, PWC и HH, но такие локальные исследования, как данное, имеет практическое применение нежелезнодорожное ознакомительное ввиду того, что дают характеристику более специфической области рынка вакансий.

Информационные технологии довольно быстро проникают во все сферы жизни общества, поэтому выпускники IT-профиля университета должны обладать широким спектром компетенций, являющимися универсальными для различных предметных областей. Одной из таких компетенций является умение анализировать статистические данные независимо от их принадлежности этим областям. Глубоко изученные и обоснованные в теории и реализованные в пакетах различных прикладных программ такие универсальные методы как дисперсионный, регрессионный, кластерный, и другие методы прикладной статистики продолжают совершенствоваться в методах глубокого машинного обучения и применяться при анализе больших данных (big data).

Для работы с данными и их анализа необходимо извлечь их из веб-ресурса. Весь этот процесс может проводиться вручную, но при большом объеме массивов данных данная задача становится невыполнимой и требует автоматизации. Здесь возникает потребность в специальных компьютерных программах, называемых парсерами или веб-скрейперами. Веб-парсер сканирует веб-страницы, загружает контент на локальный диск пользователя, извлекает из полученного массива данных необходимую информацию и затем сохраняет всё это в файлах или базе данных.

Площадка HeadHunter предоставляет специальный API, который позволяет извлекать данные из соответствующей базы данных и интегрировать HeadHunter с приложениями, разрабатываемыми сторонними разработчиками, в частности, разработчиками веб-парсеров. С помощью

API HeadHunter можно получить доступ к базам компаний на Украине, в Белоруссии, в Азербайджане, Узбекистане, Казахстане, Грузии, Киргизии. Эта возможность позволяет крупным компаниям расширять географию своего бизнеса за пределами России, а также трудоустраиваться гражданам бывшего Советского Союза на территории России.

Основные гипотезы, которые подлежали проверке:

- как различаются заработные платы специалистов от опыта работы (в годах) и уровня компетенции (junior, middle, senior);
- влияет ли язык программирования на размер заработной платы;
- специалисты какой квалификации более востребованы на рынке;
- существует ли достоверное различие между заработными платами специалистов, имеющих различные специализации.

Так рассмотрим результат проверки некоторых гипотез. Для определения факторов, оказывающих влияние на величину начального уровня заработной платы, проводился дисперсионный анализ Фишера, результат которого показал различие в среднем уровне заработной платы в зависимости предметной области и валюты.

На рисунке 1 представлены диаграммы начальных средних зарплат в различных предметных областях.

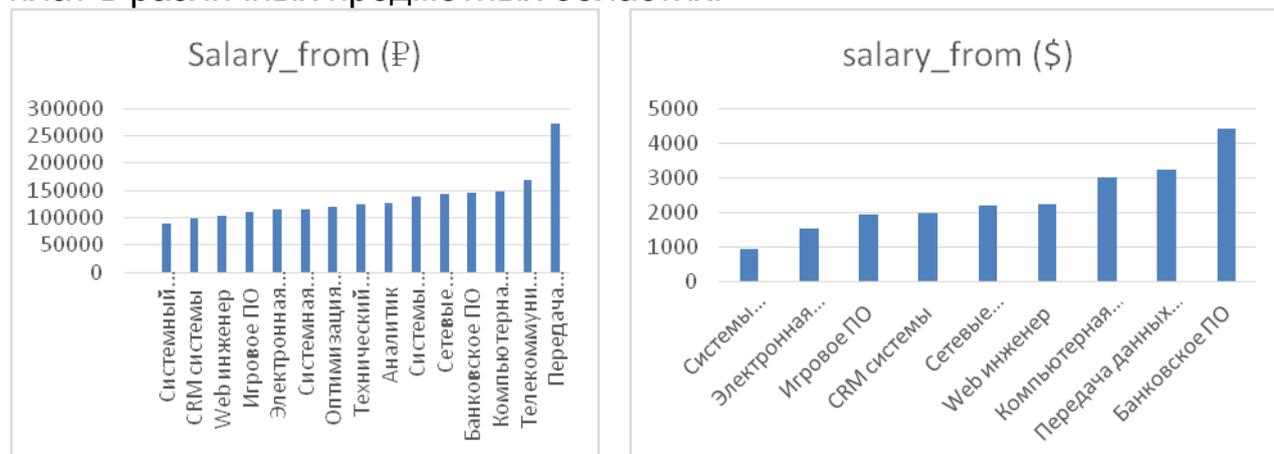


Рисунок 1 - Среднее значение начального уровня заработной платы в зависимости от предметной области

Следует отметить, что средние начальные заработные платы в таких предметных областях, как банковское ПО, передача данных и доступ в интернет, компьютерная безопасность значительно выше, чем в других сферах, и предполагают гораздо более глубокую подготовку и меру ответственности за неквалифицированную работу.

Список использованных источников:

1. Википедия – свободная энциклопедия [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/HeadHunter/>, свободный – (15.04.2021).