Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

«Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ

Школа бакалавриата

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

Руководитель образовательной программы

09.03.04 Программная инженерия

И. Н. Обабков

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

«\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2024 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

Пояснительная записка

ФОРМИРОВАНИЕ ТРЕБОВАНИЙ ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕССА ПО АКТУАЛИЗАЦИИ ЗАРАБОТНОЙ ПЛАТЫ СОТРУДНИКОВ ДЛЯ КОМПАНИИ **«**Artsofte»

|  |  |
| --- | --- |
| Руководитель: Шадрин Д. Б.  ст. преподаватель | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Нормоконтролер:В. Н. Васина | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Студент группы РИ-400015 Д. А.Старцев | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Екатеринбург

2024

**ЗАДАНИЕРЕФЕРАТ**

[**ВВЕДЕНИЕ** 5](#_Toc163252350)

[**1 Анализ предметной области** 7](#_Toc163252351)

[1.1 Проблема заказчика 7](#_Toc163252352)

[1.2 Анализ процесса формирования главного отчета 9](#_Toc163252353)

[1.3 Анализ конкурентов 13](#_Toc163252354)

[**2 Анализ средств для разработки** 15](#_Toc163252355)

[2.1 Архитектура приложения 15](#_Toc163252356)

[2.2 База данных 18](#_Toc163252357)

[2.3 Язык программирования 21](#_Toc163252358)

[**3 Техническое задание** 22](#_Toc163252359)

[3.1 Термины и определения 22](#_Toc163252360)

[**4. Проектирование и реализация** 27](#_Toc163252361)

[4.1 Обновленный процесс сбора наглядной аналитики 27](#_Toc163252362)

[4.2 Выбор лингвистической модели 28](#_Toc163252363)

[4.3 User-flow 30](#_Toc163252364)

[**ЗАКЛЮЧЕНИЕ** 33](#_Toc163252365)

[**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ** 34](#_Toc163252366)

# **ВВЕДЕНИЕ**

В современном мире цифровизация и автоматизация проникают во все сферы деятельности компаний и повседневной жизни человека. Если раньше технологии касались только технический отделов, то сегодня все нуждается в автоматизации процессов для упрощения работы и повышения ее эффективности. В том числе профессиональная HR-деятельность, без которой не обойдется ни одна компания.

Часто компании прибегают к созданию HRM (Human Resourses Management) отдела, который занимается стратегией обращения с группами сотрудников. Решения HRM должны основываться на определенных данных, сбор которых представляет собой большую работу, которая занимает много ресурсов.

В компании Artsofte В компании работает более 450 человек и оборот кадров всегда является острым вопросом для HR отдела.

Чтобы быть уверенными в своей позиции на рынке кадров: каждый квартал происходит анализ предложения на самом известном сайте по подбору вакансий — hh.ru.

Сбор и систематизация информации о вакансиях на популярном ресурсе hh.ru позволяет не только создать обширную базу данных, но и выделить ключевые тенденции в требованиях работодателей и ожиданиях соискателей. Это обеспечит эффективную адаптацию образовательных программ, обучения кадров и стратегий подбора персонала, что является актуальной задачей в условиях постоянно меняющегося рынка труда.

Чтобы уменьшить нагрузку на HR отдел и эффективно распределить его ресурс - было принято решение разработать сервис для автоматизации процесса актуализации заработной платы сотрудников.

Тем самым целью работы является аналитика процессов организации Artsofte и формирование требований к сервису автоматического анализа рынка – HRParser.

Для достижения цели поставлены следующие задачи:

1. изучение процессов организации встречи в «Artsofte»,
2. выявление и описание требований к системе у заказчика,
3. описание функциональных требование к системе,
4. описание сценариев,
5. передача технического задания в разработку.

Объектом исследования является процесс актуализации заработной платы сотрудников и возможности его автоматизации.

# **1 Анализ предметной области**

Перед тем как начать создавать продукт, необходимо провести анализ предметной области. Это включает в себя изучение бизнес-процессов компании, инструментов разработки компании, выявление и изучение проблем, общение с заказчиком, формулирование и проверку гипотез, а также разработку пользовательских историй для лучшего понимания опыта пользователя.

Эта часть работы будет фокусироваться на поиске и изучении проблемы, общении с потенциальными пользователями, и сборе требований. Затем на основе собранной информации будет составлено техническое задание, которое станет основой для разработки архитектуры, выбора инструментов и технологий разработки.

## 1.1 Проблема заказчика

Существует процесс анализа рынка труда с последующей актуализацией заработной платы сотрудников.

Сама актуализация происходит по множественным причинам:

1. Актуализация заработной платы позволяет компании прогнозировать будущие расходы на оплату труда. Это важно для предотвращения неожиданных финансовых нагрузок;
2. Регулярная актуализация заработной платы позволяет сотрудникам видеть, что их труд и профессиональное развитие оцениваются. Это способствует повышению мотивации, уровня удовлетворенности и приверженности компании;
3. Анализ рынка показывает актуальность вакансии: соотношение количества вакансий к заработной плате говорит о востребованности;
4. Знание актуальной цены кадра позволяет компании оценить свою конкурентноспособность на рынке труда. А также сравнивать условия труда.

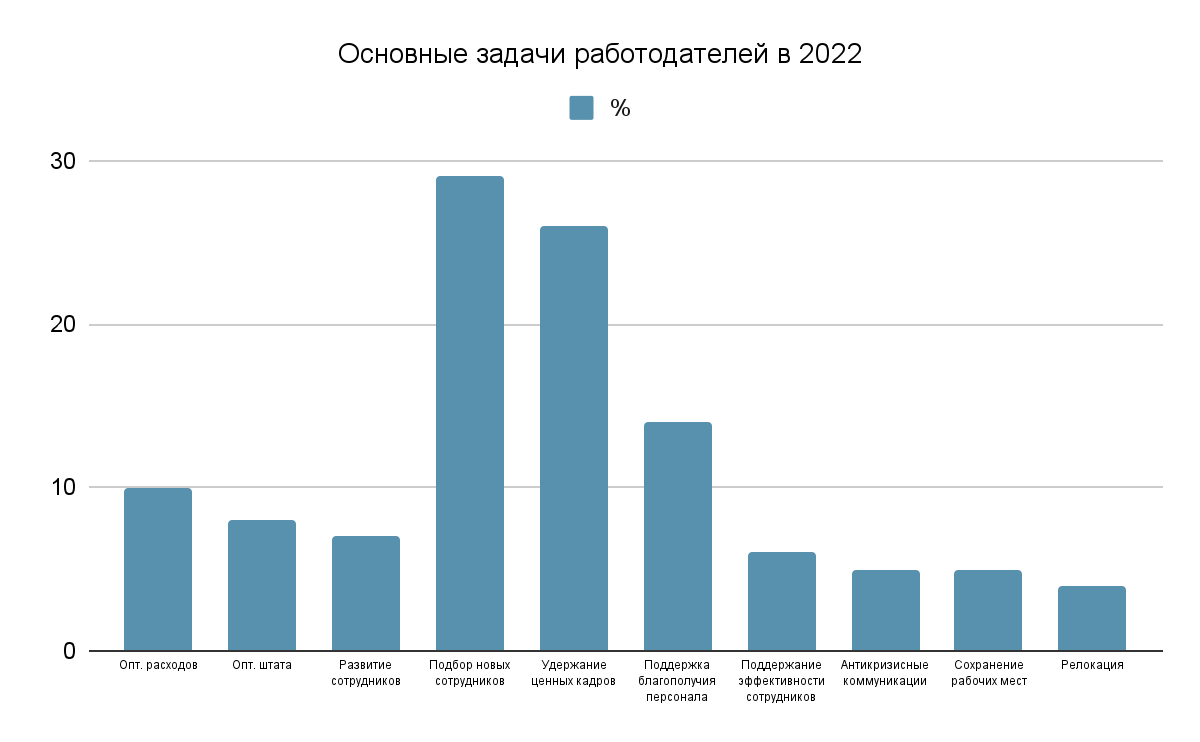


Рисунок 1 – Главные задачи работодателей в условиях кризиса по исследованию «TalentTech» в 2022 г.

На рисунке 1 приведены главные задачи работодателей в условиях кризиса по исследованию «TalentTech» в 2022 [1]

Относительно этого исследования актуализация заработной платы связана со следующими пунктами:

* Оптимизация расходов,
* Оптимизация штата,
* Подбор новых сотрудников,
* Удержание ценных кадров,
* Поддержание эффективности сотрудников,
* Сохранение рабочих мест.

То есть целевой бизнес процесс покрывает около 83% задач работодателя в условиях кризиса. Отсюда следует чуть ли не обязательное налиие этого процесса в любой IT компании современного времени. [2]

Совсем недавно рост онлайн-активности принес технологическим компаниям исторически высокие прибыли и вызвал бурное увлечение наймом сотрудников, чтобы удовлетворить растущий спрос. В этот период технологические компании предполагали, что такой взлет станет новым стандартом, что привело к активному расширению команд и быстрому увеличению размеров компаний. За время пандемии Amazon увеличил свою рабочую силу на 93%, Microsoft — на 53%, Meta — на 92%, Apple — на 20%, а Alphabet — на 60%.

У заказчика «Artsofte» имеется 120 ролей и более 450 сотрудников. Здесь под «ролью» понимается функциональная позиция в организации с определенным уровнем квалификации. Текущий бизнес процесс формирования отчета об актуальности зарплат очень затратный по ресурсам компании. Это тратит время квалифицированного сотрудника. Из-за чего приходится двигать сроки по остальным задачам. Также если ускорить весь процесс, то компания будет иметь возможность чаще актуализировать заработную плату сотрудникам. Сейчас это происходит раз в полугодие.

Чтобы решить эту проблему нужно найти участок бизнес процесса, который является самым трудозатратным и автоматизировать работу на этом участке.

## 1.2 Анализ процесса формирования главного отчета

Удобно будет описать процесс формирования итогового отчета с результирующей заработной платой через Customer Touchpoint Map (рисунок 2) – это визуальное представление всех важных моментов взаимодействия клиента с брендом или продуктом на протяжении всего пути от первого знакомства до последующих этапов, целью которой является оптимизация клиентского опыта.

A diagram of a project

Description automatically generated with medium confidence

Рисунок 2 – Customer Touchpoint Map процесса создания итового отчета

Процесс делится на 5 составляющих:

1. Сбор обезличенной аналитики,
2. Сбор наглядной аналитики,
3. Анализ собранных данных,
4. Формирование итоговой презентации,
5. Презентация перед руководством.

Сбор обезличенной аналитики – на этом этапе собираются данные по зарплатам из открытых источников. Данные в таких метриках уже структурированы. На данный момент лучшими из бесплатных решений являются:

* <https://career.habr.com/salaries>
* <https://rit.work/jobs>

Отчеты со сторонних источников, хотя и предоставляют обширные обзоры, вносят свои собственные ограничения в виде ограниченного объема данных и отсутствия возможности проверки их достоверности. Несмотря на эти ограничения, собранный датасет играет важную роль в процессе отвалидации аналитики на следующем этапе исследования.

На основе анализа этих обезличенных данных мы можем выделить ключевые тенденции и закономерности, которые будут служить основой для более глубоких исследований на следующих этапах. Важно подчеркнуть, что, несмотря на отсутствие возможности проверки конкретных вакансий, полученные данные предоставляют ценный обзор зарплатного уровня в индустрии и являются неотъемлемой частью проверочного этапа нашего исследования.

Стоит отдельно отметить, что hr не контролирует метрики, которые ему предоставляют сторонние ресурсы по анализу рынка, из-за чего тратятся дополнительные ресурсы на анализ полученных датасетов.

Сбор наглядной аналитики – на этом этапе формируется отчет похожий на тот, что на предыдущем шаге, однако в этом отчете мы можем сами настроить нужные нам метрики, брать только нужные нам данные. Отчет формируется вручную по каждой роли.

A diagram of a flowchart

Description automatically generated

Рисунок 3 – BPMN диаграмма процесса «Сбор наглядной аналитики»

По схеме процесса «Сбор данных для наглядной аналитики» (рисунок 3) можно понять, что работа распараллеливается на нескольких человек из-за большого объема ручной работы.

Работа каждого юнита (схема Сбор данных в рамках роли) заключается в том, что он заходит на сайт hh.ru, начинает читать вакансии и механически переносить данные со страницы вакансий в excel таблицу – построчно формируя отчет по одной роли. На это уходит большинство времени.

A graph with blue and orange bars

Description automatically generated

Рисунок 4 – Распределение времени среди трёх этапов создания итогового отчёта

Как можно заметить по рисунку 4 – сбор данных для этапа «Наглядная аналитика #2» занимает около 74% от суммарного времени всех этапов. Именно здесь происходит «утечка» времени. Следвательно именно этот этап и стоит подвергать автоматизации.

Анализ собранных данных представляет собой выявление необходимых для итогового отчета данных:

* Зарплатные ожидания
* Анализ уровня конкурентоспособности
* Тенденции рынка
* Адаптация стратегии найма
* Оптимизация зарплат внутри компании

Формирование итоговой презентации – здесь данные, полученные на предыдущем этапе представляются в упорядоченном виде для упрощения восприятия аналитических данных. Здесь происходит верстка презентации, размещение смысловых акцентов. Этот этап нужен для подведения итогов аналитической работы.

Презентация перед руководством – на этом этапе отчет презентуют пред руководством, которое будет принимать решение по оптимизации бюджета на следующее полугодие.

Таким образом, для автоматизации процесса актуализации заработной платы сотрудников необходимо: разработать систему, которая автоматически собирает данные с hh.ru и формирует отчет по запросу пользователя. Тем самым сократив время всего процесса примерно на две трети.

## 1.3 Анализ конкурентов

Главным приемуществом нашего продукта по сравнению с конкурентами следует считать возможность формировать специфические метрики. Далее приводится список ресурсов, которые уже используются для анализа рынка. Также приводится список их приемуществ и недостатков.

* Habr Карьера [3],
* HeadHunter [4],
* Rit.work [5],
* Роснавык [6],
* Зарплатомер (SuperJob) [7].

Habr карьера:

Источник данных: система получения данных на Habr-карьера заключается в следующем. Пользователь, который хочет узнать ситуацию на рынке труда – должен указать свой уровень заработной платы и должность (профессию). Таким образом пользователь попадает в общий набор данных, который и предоставляется к анализу на этом портале.

То есть – ничто не мешает пользователю внести случайные данные, проверить это не выйдет, т.к. хабр предоставляет только итоговые данных, которые являются обезличенными.

Итоговая информация представляет собой набор квалификаций (Lead, Senior, Middle, Junior, Intern) – и соответствующее распределение зарплат. Также имеется возможность фильтрации по местоположению, специализации, профессиональным навыкам, компаниям, типу занятости и формату работы.

HeadHunter:

Самый популярный сайт для работодателей и соискателей в СНГ. Источником данных являются реальные заказчики рабочей силы. Каждую вакансию можно проверить на подлинность.

Итоговой информацией здесь является набор вакансий, которые можно свободно фильтровать по различным критериям.

Его недостаток лишь в том, что данные представлены как есть, то есть в слабо упорядоченном виде и наложить дополнительный фильтр на данные или собрать метрику (например медиану по выборке) – нельзя. Отчего нужно создавать свои инструменты анализа по данным, которые HeadHunter предоставляет.

Rit.work:

Это HeadHunter, но только для IT-вакансий. Здесь своя база данных, преимущество над HeadHunter`ом в том, что помимо тех же фильтров, что и на вышеупомянутой платформе, есть вывод метрики по отфильтрованным данным (перцентили, график - зарплата по месяцам, график - количество/зарплата).

Количество вакансий по сравнению с ХХ заметно меньше и обновляются они менее динамично, чем на самом популярном поисковике СНГ.

Роснавык:

Хорошая визуализация данных, удобные отчеты. Однако нельзя «провалиться» в вакансию и узнать её источник. Ограниченное количество фильтров для одного запроса.

Также есть платные аналитические отчеты по согласованному списку профессиональных областей (до 4ех штук) по 50000 рублей за одну штуку.

Источники информации – Работа России, HeadHunter, SuperJob, Работа.ру. Проверки на дубли не ведется, отсюда можно получить погрешности при анализе данных.

SuperJob:

Удобные полезные отчеты по запрошенной профессии. Обезличенные данные. Высокая цена.

Результатом данного анализа можно представить в виде таблицы

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сервис | Обезличенность | Стоимость | Метрики | Наполнение |
| Habr Карьера | 1 | 5 | 3 | 2 |
| HeadHunter | 5 | 5 | 1 | 5 |
| Rit.work | 5 | 5 | 3 | 2 |
| Роснавык | 1 | 3 | 5 | 3 |
| SuperJob | 1 | 1 | 4 | 5 |
| Парсер | 5 | 4 | 5 | 5 |

Таблица 1 Сравнительный анализ конкурентов

# **2 Анализ средств для разработки**

## 2.1 Архитектура приложения

Программный продукт представляет собой два сервиса, связанных друг с другом технологией REST – это стиль архитектуры программного обеспечения для распределенных систем, ключевой особенностью которого является отсутствие состояния (statelessness). Это означает, что каждый запрос от клиента к серверу должен содержать всю необходимую информацию для его выполнения, без использования данных о предыдущих запросах.

A black and white screen with white text

Description automatically generated

Рисунок 5 – Архитектура взаимодействия сервисов

На рисунке 5 изображениа схема двух сервисов, которые и являют собой систему обработки вакансий с hh.ru. И обеспечивают вывод информации в виде xlsx документов.

В рамках данного продукта моя зона ответственности – клиентская сторона личного кабинета пользователя (lk–user). Иными словами – верстка шаблонов, дизайн интерфейсов, добавление динамики при помощи средств JavaScript, внедрение JavaScript файлов в монолит при помощи фреймворка Webpack для соблюдения принципов чистого кода. В этом сервисе не представлено никакой сложной логики. Просто набор интерфейсов и база данных для хранения сопутствующих сущностей.

Второй сервис (сервис аналитики – lk–parser) – представляет собой парсер, генератор отчетов, калькулятор аналитической информации, прокси запросов к лингвистической модели и также хранилище данных для аналитики.

Взаимодействие между сервисами осуществляется посредством протокола HTTP, где необходимая информация передается в формате JSON, что соответствует принципам архитектуры REST. Каждый сервис оперирует собственным набором баз данных, обеспечивая тем самым высокую степень автономности. Это означает, что в случае неполадок или отказа одного из сервисов, другие могут продолжать свою работу без существенного влияния на общую функциональность системы.

Redis, представляющая собой NoSQL базу данных, служит для эффективного управления кэшем данных в распределенных средах. Она обеспечивает высокую производительность за счет хранения данных в оперативной памяти и использования сложных алгоритмов кэширования. Это позволяет уменьшить нагрузку на основные базы данных и сократить временные задержки при обращении к данным.

PostgreSQL, в свою очередь, является реляционной базой данных, предназначенной для долгосрочного и структурированного хранения данных. Ее гибкий и мощный SQL-движок обеспечивает эффективное хранение и обработку данных в различных сценариях, от малых приложений до крупных корпоративных систем.

RabbitMQ, являясь брокером сообщений, выполняет роль посредника между различными компонентами системы, обеспечивая надежную и отказоустойчивую передачу сообщений. Его механизмы хранения сообщений гарантируют сохранность данных даже в случае сбоев или отказов в работе сервисов, обеспечивая непрерывность работы системы и сохранение целостности данных.Каждый сервис находится в изолированном Docker контейнере, что позволяет обеспечить максимальную безопасность и устойчивость системы. Контейнеризация с помощью Docker обеспечивает изоляцию каждого сервиса, предотвращая взаимное воздействие между ними и минимизируя возможность уязвимостей. Кроме того, такой подход упрощает развертывание, масштабирование и управление сервисами, делая процесс более эффективным и прозрачным.

## 2.2 База данных

A screenshot of a computer

Description automatically generated

Рисунок 6 – База данных сервиса личных кабинетов

На рисунке 6 представлена база данных сервиса lk–user, ниже представлено её описание.

База данных сервиса личного кабинета состоит из 6 сущностей:

Пользователь – чтобы обеспечивать ограниченный доступ к ресурсу

Миграции – для удобства развертывания приложения и актуализации базы данных при разработке

Hh-регион – регионы хранятся в базе данных чтобы не запрашивать их с hh при каждом формировании аналитического запроса

Hh-индустрия – аналогично региону

Hh-подиндустрия – аналогично региону, связана с hh-индустрией первичным ключём

Аналитический запрос – аналитический запрос хранится на стороне личного кабинета чтобы можно было переотравлять запросы при неудачном выполнении.

На рисунке 7 представлена база данных сервиса lk–parser, ниже представлено её описание.

База данных сервиса аналитики опирается на 5 сущностей (плюс 3 вспомогательных).

Сущность вакансии (vacancy) – аналитическая единица отчета, именно эта сущность обогащается данными при парсинге и дополнительной обработке.

Также в отдельную таблицу (order\_excel\_file\_param) записываются параметры формируемых файлов для более удобного обращения к ним при скачивании.

Formula\_setting и calculation\_results служат для сохранения применяемых к вакансии формул при обработке.

Также существует таблица, дублирующая таблицу из сервиса личного кабинета – parser\_order. Это сделано для репликации данных, чтобы в случае чего не было потерь важной информации.

A screenshot of a computer

Description automatically generated

Рисунок 7 – База данных сервиса аналитики

## 2.3 Язык программирования

Было принято решение разрабатывать сервис личных кабинетов связкой языков JavaScript и PHP. В частности, фреймворков jQuery, Symfony.

Выбор использования связки JavaScript (JQuery) + PHP Symfony с интеграцией Webpack и Bootstrap обусловлен несколькими ключевыми факторами, которые обеспечивают эффективное производство и достижение целей сервиса:

Производительность и нагрузка: Учитывая, что сервис не испытывает высокой нагрузки и выполняет относительно простые задачи, выбор легковесных технологий таких как JavaScript (JQuery) и PHP Symfony позволяет достичь необходимого уровня производительности без излишней сложности и затрат.

Экономичность: Использование открытых и бесплатных технологий, таких как JavaScript (JQuery) и PHP Symfony, а также фреймворка Bootstrap, помогает сократить расходы на лицензирование и разработку. Это особенно важно, учитывая дешевизну производства.

Удобство разработки и поддержки: Symfony предоставляет мощные инструменты для разработки веб-приложений, обеспечивая высокую производительность и удобство в работе с базой данных и другими аспектами приложения. Интеграция с Webpack позволяет управлять зависимостями JavaScript и оптимизировать их загрузку, а Bootstrap предоставляет готовые компоненты и стили для быстрой разработки пользовательского интерфейса.

Гибкость и масштабируемость: Даже если в данный момент сервис не испытывает высокой нагрузки, выбранный стек технологий обеспечивает достаточную гибкость и масштабируемость для возможного расширения функциональности и увеличения нагрузки в будущем без необходимости пересмотра архитектуры приложения.

Таким образом, использование связки JavaScript (JQuery) + PHP Symfony с интеграцией Webpack и Bootstrap обеспечивает оптимальное сочетание экономичности, производительности, удобства разработки и гибкости, соответствуя потребностям и возможностям нашего сервиса.

# **3 Техническое задание**

Техническое задание предоставлено в свободной форме и представляет собой фиксацию ключевых для разработки документов: содержание xlsx файла отчета, контекст для запроса в лингвистическую модель, содержание аналитического запроса.

## 3.1 Термины и определения

|  |  |
| --- | --- |
| Термин /  Сокращение | Определение / Расшифровка |
| Отчет | Файл в формате xlsx, содержащий в структурированном виде данные по Аналитическому запросу |
| Аналитический запрос | Запрос в модуль парсера по которому формируется Отчет и Страница отчета. (формируется на странице запроса) |
| Страница отчета | Web страница размещенная в личном кабинете пользователя, содержащая в структурированном виде данные по Аналитическому запросу |
| Личный кабинет пользователя | Модуль системы, отвечающий за доступ в хранилище отчетов и страниц отчета |
| Парсер | Модуль системы отвечающий за сбор данных по вакансиям |

Таблица 2 Термины и определения технического задания

В таблице 2 перечислены ключевые термины для последующего понимания содержания технической документации.

3.2 Требования к данным

|  |  |  |  |
| --- | --- | --- | --- |
| Столбец | Пояснение | Определяется нейросетью | Пример |
| Ссылка на вакансию в hh.ru | Прямая ссылка на страницу вакансии на hh.ru | - | https://ekaterinburg.hh.ru/vacancy/90947747 |
| Набор навыков из словаря hh.ru | Ключевые навыки, которые работодатель ожидает у соискателя, взятые из словаря hh.ru | + | Python, SQL, JavaScript, аналитическое мышление, коммуникабельность |
| Название вакансии | Здесь указывается официальное название вакансии, как оно представлено на площадке hh.ru | - | Специалист по разработке программного обеспечения |
| Грейд | Уровень квалификации указанный в вакансии | + | Junior |
| Задача и функционал | Описание основных задач и обязанностей, которые предстоит выполнять сотруднику в данной вакансии | + | Разработка и поддержка программного обеспечения, участие в проектировании систем, тестирование и оптимизация кода |
| Описание вакансии | Описание самой вакансии, предоставленное работодателем | - | Мы ищем опытного специалиста с углубленными знаниями в области программирования для участия в проектах по созданию высокотехнологичных решений |
| Требования в вакансии | Опыт работы | - | Опыт работы от 3 лет, знание Python и опыт работы с базами данных, высшее техническое образование |
| Наличие ДМС | Информация о наличии медицинской страховки | + | Да |
| Рабочее время | Информация о наличии рабочем графике | - | Гибкий график |
| Удаленная работа | Информация о возможности удаленной работы | - | Да |
| Результирующий уровень заработной платы | Диапазон заработной платы, который предоставляется для данной вакансии | - | 80 000 - 120 000 |
| Название компании | Название компании, предлагающей вакансию | - | ArtSofte |
| ID компании | ID компании с сайта hh.ru | - | 1234567 |
| Дополнительная информация по заработной плате | | | |
| Медиана по выборке | | | |
| Все моды по выборке | | | |
| Средняя по выборке | | | |
| Перцентили по 25% | | | |

Таблица 3 – содержание xlsx файла (отчета)

В таблице 3 содержится информация, которая соответствует одной вакансии в итоговом отчете, который представляет собой xlsx файл.

Контекст запроса к лингвистической модели:

Ты получил на вход заголовок вакансии и описание вакансии.

Тебе нужно дать ответ в JSON и только в виде JSON вида:

{

“VMI”: true/false,

“functional”: “some\_functional\_by\_description”,

“keySkills”: {“skill1”, “skill2”, “skill3”},

“grade”: “enum: junior, middle, senior, teamlead”

},

где VMI – наличие ДМС в вакансии, functional – чем предстоит заниматься на рабочем месте (функиональность вакансии), keySkills – набор ключевых навыков из вакансии в виде массива, grade – опередли уровень квалификации сотрудника по описанию вакансии и подставь нужный уровень из списка: junior, middle, senior, teamlead.

ОЧЕНЬ ВАЖНО! Если заголовок не соответствует описанию вакансии верни ответ в виде JSON вида:

{

“success” : false

}

И ничего больше.

|  |  |
| --- | --- |
| Название поля | Предназначение |
| Текст поиска | Аналогичен «умному поиску» с hh.ru (описание есть в <https://ekaterinburg.hh.ru/article/1175>) |
| Уровень квалификации сотрудника | Поиск производится только среди этого уровня квалификации |
| Регион поиска | Позволяет получить выборку вакансий по определённому региону |
| Наличие ДМС | Предоставляет вакансии только с наличием ДМС |
| Поля поиска | Модифицирует поле «Текст поиска», предоставляя опции: Поиск в названии  Поиск в названии компании  Поиск в описании вакансии |
| Список индустрий | Возможность отфильтровать результат запроса по списку индустрий hh.ru |
| Вакансии только с заработной платой | При включенной опции вакансии без указания заработной платы будут исключены из отчета |
| Тип занятости | Возможность отфильтровать результат запроса по полям:  Полная занятость  Частичная занятость  Проектная работа/разовое задание  Волонтерство  Стажировка  Оформление по ГПХ или по совместительству |
| График работы | Возможность отфильтровать результат запроса по полям:  Полный день  Сменный график  Гибкий график  Удаленная работа  Вахтовый метод |
| Опыт работы | Возможность отфильтровать результат запроса по полям:  Не имеет значения  Нет опыта  От 1 года до 3 лет  От 3 до 6 лет  Более 6 лет |

Таблица 4 – аналитический запрос

В таблице 4 представлен запрос, который поступает в сервис аналитики из формы запроса сервиса личных кабинетов

# **4. Проектирование и реализация**

## 4.1 Обновленный процесс сбора наглядной аналитики

Существующий процесс представлен на рисунке 3. Его подробное описание выглядит так: Собираем список должностей для анализа. Обычно этих ролей около 1, в связи с этим работа по обработке разделяется на несколько человек. Каждый HR берет из «стопки должностей» одну и формирует по ней xlsx файл: сначала по названию должности подается запрос в hh.ru, чтобы найти актуалные вакансии по этому названию. Каждая вакансия соответствует строке xlsx файла. Сама строка формируется путем сбора следующей информации из опиания вакансии:

1. Извлечение задач и функционала,
2. Извлечение ифнормации о дополнительных условиях труда,
3. Извлечение результирующего уровня заработной платы,
4. Извлечение дополных данных для верстки отчета.

После того, как по каждой вакансии сформирована строка – xlsx файл считается законченым. И берутся следующие должности до тех пор, пока они не закончатся.

Тогда определим слабые стороны старого процесса

* Над обработкой ролей работает несколько человек,
* Каждая отдельная вакансия обрабатывается в несколько сложных этапов,
* Основное время уходит на рутинные действия, которые легко автоматизировать.

Следовательно в новом процессе эти недостатки должны быть исправлены. Тогда новый процесс должен выглядеть следующим образом:

A diagram of a process

Description automatically generated

Рисунок 8 – обновленный процесс получения отчета

Теперь процесс выглядит так: получаем выгрузку должностей из битрикса, выполняем запрос на формирования отчета по каждой должности. Далее запросы попадают в сервис формирования отчетов. (Там, по определенному запросу в LLM, автоматизируются ручные действия по вычленению данных из текста.) Ожидаем выполнения всех запросов. Данные собраны и представлены в xlsx формате.

## 4.2 Выбор лингвистической модели

Важной частью автоматизации является нейросеть по вычленению слов по заданному контексту. Есть много различных решений этой конкретной задачи. Нужно выбрать ту, которая устроила бы заказчика и была легко интегрируема в наш сервис.

Есть два подхода к реализации работы нейросети в сервисе:

* Облачное решение,
* Компиляция кода нейросети.

Компилирование нейросети ведет к высокой ресурсной нагрузке. Т.к. вычисления просиходят на графическом ядре. Отсюда следуют высокие расходы на инфраструктуру проекта. (В сравнении со стоимостью облачного решения)

Помимо поддержки инфраструктуры, для скомпилированной нейросети нужно иметь большой датасет для тренировки этой самой модели.

количество примеров в датасете должно быть около 50тыс единиц для более-менее валидной работы этой самой нейросети. Чем больше - тем лучше. Обработка одной вакансии занимает 1-2 минуты. Таким образом сбор датасета займет около 1250 часов.

Также, натренированная нейросеть не учитывает свободный словесный запрос по типу: «если в твоей выборке будет вакансия на тимлида, то не включай эту вакансию в выборку, просто исключи её.»

Нейросеть, которая не является большой лингвистической моделью - не способна воспринимать такие просьбы пользователя.

Также использование ChatGPT обусловлено дешевизной разработки, то есть это уже готовое решение, которое свободно интегрируется посредством открытого API.

Т.к. прямые запросы в ChatGPT из России невозможны - нужно использовать proxy для запросов. Таким образом выбор пал на neuroapi, где стоимость запроса в 500 токенов (примерно столько занимает запрос одной вакансии) равна 2коп. То есть при среднем объеме отчета в 600-1000 вакансий (запросов) и 120 должностях. Получаем расход соответствующий диапазону 600 \* 120 \* 0.02 <= x <= 1000 \* 120 \* 0.02 или 1440 <= x <= 2400 рублей за одну реализацию процесса по актуализации заработной платы сотрудников «Artsofte»

Сервер с видеокартой для реализации LLM нейросети (или же упрощенной NLP - Natural Language Processing) стоит в месяц минимум 7500рублей (самый дешевый тариф на hostkey.ru). Есть возможность брать сервера посуточно (около 400 рублей в сутки, что примерно равно использованию прокси), но тогда все запросы нужно будет формировать заранее и исключается возможнось «экстренного парсинга», Также это накладывает лишние расходы на тестирование и разработку продукта.

## 4.3 User-flow

Для демонстрации работы с продуктом предлагаю ознакомится с user-flow решения (рисунок 9). User flow (поток пользователя) представляет собой последовательность действий, которые пользователь выполняет при взаимодействии с продуктом или сервисом. Это концептуальная модель, которая описывает путь пользователя от начальной точки (например, посещение веб-сайта или запуск приложения) до достижения конечной цели (например, совершение покупки, регистрация аккаунта и т. д.).

Изображение выглядит как диаграмма, снимок экрана, линия, План

Автоматически созданное описание

Рисунок 9 – user flow получения отчета

Принцип работы с нашим решением довольно прост и не требует дополнительного обучения персонала: чтобы получить отчет необходимо войти в личный кабинет, используя логин и пароль, далее нужно зайти во вкладку Request (рисунок 10) и заполнить нужные для запроса поля.

A screenshot of a computer

Description automatically generated

Рисунок 10 – форма запроса

Запрос описан в таблице 4. Стоит отметить, что в каждом поле с мультиселектом реализован поиск по подстроке и добавлены кнопки для выбора всех вариантов. Также из-за большого списка индустрий с hh.ru было принято решения добавить кнопку для их сокрытия чтобы не перегружать интерфейс.

После того, как запрос был отправлен на сервер, можно зайти во вкладку Documents (рисунок 11) и увидеть статус документа (

START\_PARSING – парсинг начат,

COMPLETE – парсинг окончен,

GENERATE\_DOCUMENT – генерация документа,

COMPLETE\_WITH\_DOCUMENT – генерация документа окончена,

PARSING\_ERROR – ошибка при выполнении

)

A screenshot of a computer

Description automatically generated

Рисунок 11 – страница со списком отчетов

Также при нажатии на хэш можно посмотреть запрос детальнее (рисунок 12)

A screenshot of a computer

Description automatically generated

Рисунок 12 – страница с содеранием запроса

Таким образом user-flow укладывается в три интуитивных интерфейса и не требует привлечения дополнительных ресурсов на обучение персонала. Главным преимуществом этого порядка действий является то, что не нужно дожидаться выполнения одного запроса, чтобы начинать формировать следующий.

# **ЗАКЛЮЧЕНИЕ**

Разработка и внедрение системы HRParser позволило компании Artsofte решить проблему актуализации заработной платы сотрудников, оптимизировать расходы и повысить эффективность работы HR отдела. Автоматизация данного процесса стала ключевым фактором для улучшения управления персоналом и обеспечения конкурентоспособности компании на рынке труда

Анализ предметной области и выявление проблемы позволили сфокусироваться на разработке инновационного решения, которое значительно упростило и ускорило процесс актуализации заработной платы. Теперь HR отдел имеет возможность чаще проводить анализ рынка труда и обновлять данные о заработной плате сотрудников, что способствует более точному прогнозированию расходов компании и повышению мотивации персонала.

Внедрение HRParser открывает новые перспективы для компании Artsofte, позволяя ей оперативно реагировать на изменения на рынке труда, улучшать условия работы сотрудников и обеспечивать устойчивое развитие в условиях динамичного бизнес-окружения.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

* 1. HR-шторм: реакция российских работодателей на новые условия. Исследование IT-холдинга TalentTech. – URL: https://potok.io/wp-content/uploads/2022/04/issledovanie-talenttech-hr-shtorm-2022.pdf (дата обращения: 05.04.24)
  2. Исследование TalentTech: HR-шторм — 2022. Как российские работодатели отреагировали на кризис URL: https://potok.io/blog/hr-research/hr-storm-2022-russia-crisis-research/ (дата обращения 05.04.24).
  3. Информация по конкурентному сервису URL: https://career.habr.com (дата обращения 05.04.24).
  4. Информация по конкурентному сервису URL: https://hh.ru (дата обращения 05.04.24).
  5. Информация по конкурентному сервису URL: https://rit.work (дата обращения 05.04.24).
  6. Информация по конкурентному сервису URL: https://lk.rosnavyk.ru (дата обращения 05.04.24).
  7. Информация по конкурентному сервису URL: https://www.zarplatomer.ru (дата обращения 05.04.24).
  8. Горохов Е.Р., Горохова О.Ю. Бизнес-аналитика: учебник. – Москва: Финансы и статистика, 2020.
  9. Перепелкин М. В. Бизнес-аналитика: методы и технологии. – Москва: Издательство Дело, 2019.
  10. Коваленко М. Ю. Бизнес-аналитика. Основы, технологии, методы. – Санкт-Петербург: Питер, 2019.
  11. Бизнес-аналитика и системный анализ. – Москва: Издательский дом «Статистика», 2020.
  12. Проничкина Е. В. Бизнес-аналитика: современные методы и подходы. – Москва: Издательский дом «Дело», 2019.
  13. Антипин А.Е. Бизнес-аналитика в программном обеспечении. – Москва: Питер, 2019.
  14. Бизнес-аналитика и управление проектами. – Москва: Питер, 2019.
  15. Савицкая М. А. Бизнес-аналитика: фундаментальные принципы и методологии. – Москва: Издательский дом «Дело», 2019.