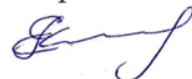


Министерство науки и высшего образования Российской Федерации
ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»
Кафедра «школа бакалавриата (школа)»

Оценка работы отлично (99)
Руководитель от УрФУ Корнякова Е.М.



Тема задания на практику

Проектирование и разработка ПО

ОТЧЕТ

Вид практики Производственная практика

Тип практики Производственная практика, преддипломная

Руководитель практики от предприятия (организации) Абросимов М.П.
ФИО руководителя Подпись



Студент Старцев Д.А.
ФИО студента



Специальность (направление подготовки) 09.03.04 Программная инженерия

Группа РИ-400015

Екатеринбург 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Анализ предметной области	5
1.1 Проблема заказчика	5
1.2 Анализ процесса формирования главного отчета	7
1.3 Анализ конкурентов.....	11
1.4 Календарный план	13
2. Практическая часть	14
2.1 Архитектура приложения.....	14
2.2 База данных.....	16
2.3 Язык программирования	17
2.4 Техническое задание.....	19
2.5 Обновленный процесс сбора наглядной аналитики	21
2.6 Выбор лингвистической модели	23
2.7 User–flow	24
ЗАКЛЮЧЕНИЕ	28
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	29
ПРИЛОЖЕНИЕ А (обязательное) Схема базы данных сервиса ЛК	31
ПРИЛОЖЕНИЕ Б (обязательное) Схема базы данных сервиса аналитики ...	32
ПРИЛОЖЕНИЕ В (обязательное) Содержание xlsx файла (отчета)	33
ПРИЛОЖЕНИЕ Г (обязательное) Аналитический запрос	37

ВВЕДЕНИЕ

В современном мире цифровизация и автоматизация проникают во все сферы деятельности компаний и повседневной жизни человека. Если раньше технологии касались только технического отдела, 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 Resources 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 «Artsoft»;
- 2) выявление и описание требований к системе у заказчика;
- 3) описание функциональных требований 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

1.1 Проблема заказчика

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.

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

Сама актуализация происходит по множественным причинам:

а) актуализация заработной платы позволяет компании прогнозировать будущие расходы на оплату труда. Это важно для предотвращения неожиданных финансовых нагрузок;

б) регулярная актуализация заработной платы позволяет сотрудникам видеть, что их труд и профессиональное развитие оцениваются. Это способствует повышению мотивации, уровня удовлетворенности и приверженности компании;

в) анализ рынка показывает актуальность вакансии: соотношение количества вакансий к заработной плате говорит о востребованности;

г) знание актуальной цены кадра позволяет компании оценить свою конкурентоспособность на рынке труда. А также сравнивать условия труда.

На рисунке 1 приведены главные задачи работодателей в условиях кризиса. График составлен на основе компиляции данных из исследования «TalentTech» в 2022 [1].

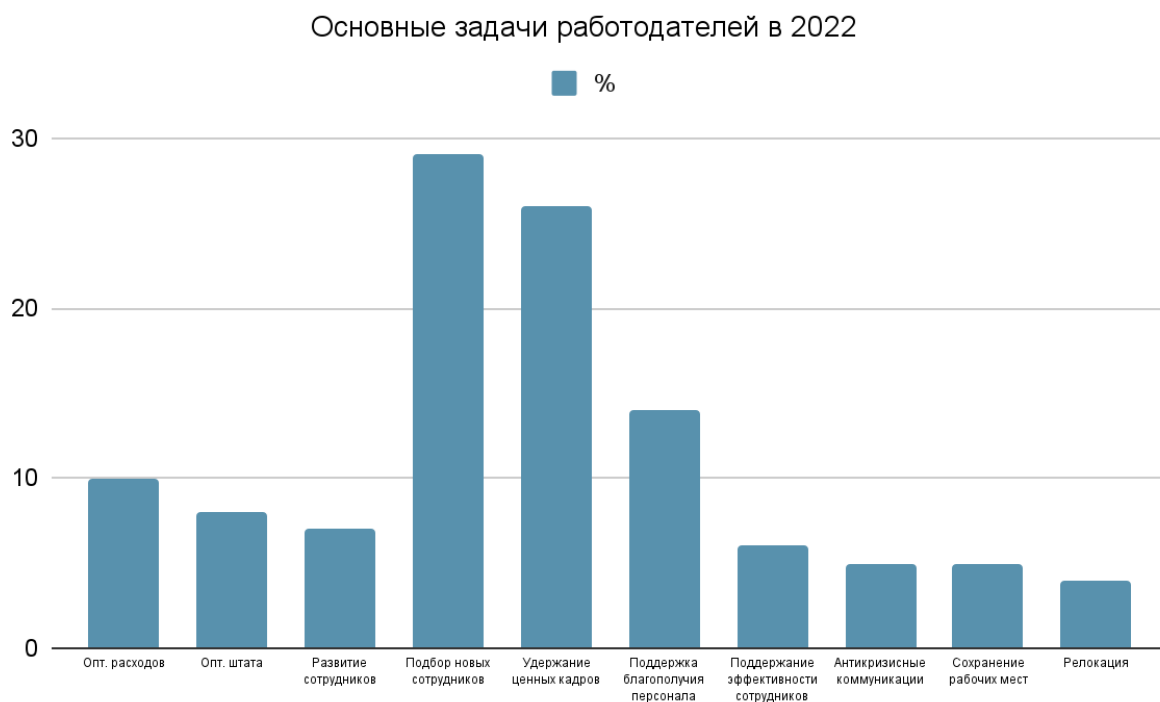


Рисунок 1 – Главные задачи работодателей в условиях кризиса по исследованию «TalentTech» в 2022 г.

Относительно этого исследования актуализация заработной платы связана со следующими пунктами:

- а) оптимизация расходов;
- б) оптимизация штата;
- в) подбор новых сотрудников;
- г) удержание ценных кадров;
- д) поддержание эффективности сотрудников;
- е) сохранение рабочих мест.

То есть целевой бизнес–процесс покрывает около 83% задач работодателя в условиях кризиса. Тогда следует чуть ли не обязательное наличие этого процесса в любой IT компании современного времени.

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. По статистике из исследования М.Тех [2] 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. Приходится переносить сроки выполнения остальных задач, в связи с этим. 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.

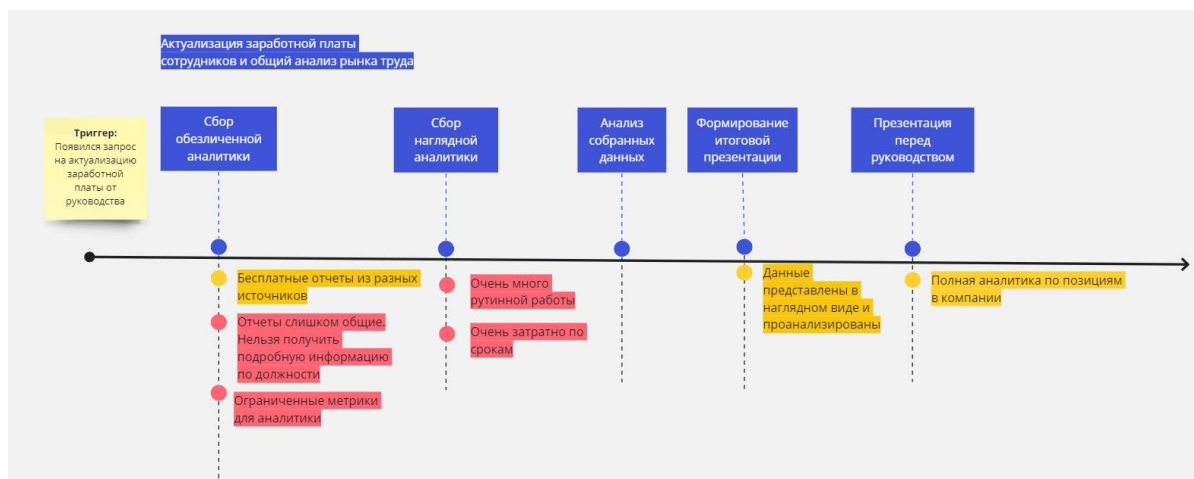


Рисунок 2 – Customer Touchpoint Map процесса создания итогового отчета

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 валидации 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, в связи с чем 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.

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

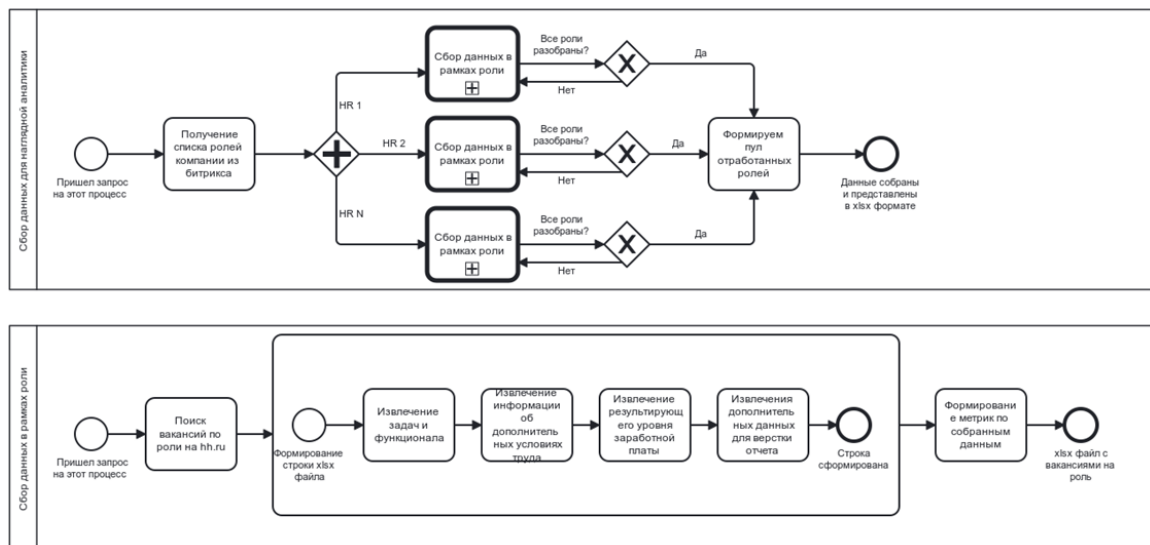


Рисунок 3 – BPMN диаграмма процесса «Сбор наглядной аналитики»

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.

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. Следовательно, именно этот этап и стоит подвергать автоматизации.

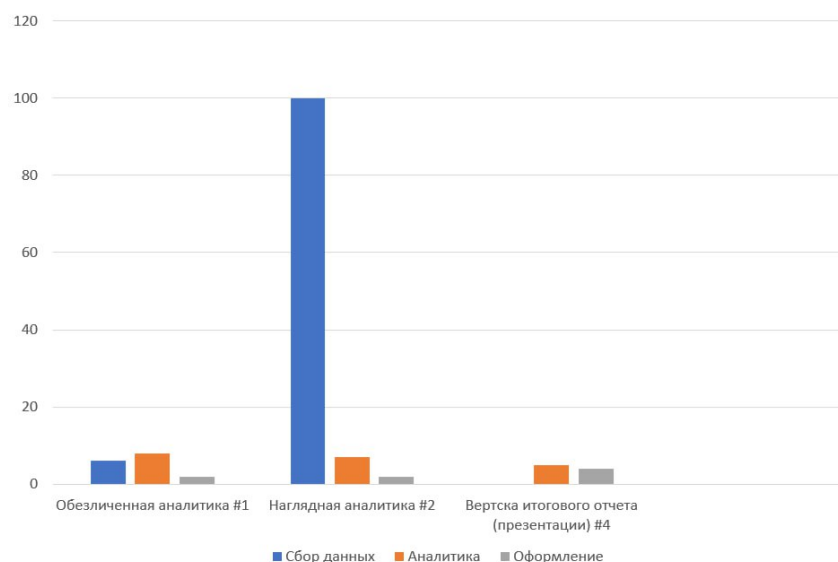


Рисунок 4 – Распределение времени среди трёх этапов создания итогового отчёта

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

- а) зарплатные ожидания;
- б) анализ уровня конкурентоспособности;
- в) тенденции рынка;
- г) адаптация стратегии найма;
- д) 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 перед 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 преимуществом 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 преимуществ и недостатков.

- а) Habr Карьера – электронный ресурс [3];
- б) HeadHunter – электронный ресурс [4];
- в) Rit.work – электронный ресурс [5];
- г) Роснавык – электронный ресурс [6];
- д) Зарплатомер (SuperJob) – электронный ресурс [7].

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 данные, 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 обладает хорошей визуализацией данных и удобными отчетами. 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.

Результаты данного анализа можно представить в виде таблицы 1, где Парсер – название нашего сервиса.

Таблица 1 – Сравнительный анализ конкурентов

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

1.4 Календарный план

Основные этапы практики приведены в таблице 2.

Таблица 2 – Этапы практики

Этап практики	Дата
Анализ бизнес–процессов компании	17.02 – 24.02
Анализ требований к системе	25.02 – 29.02
Описание технического задания и построение архитектуры системы	29.02 – 24.03
Предлагаемое техническое решение	24.03 – 05.04

2. Практическая часть

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.

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

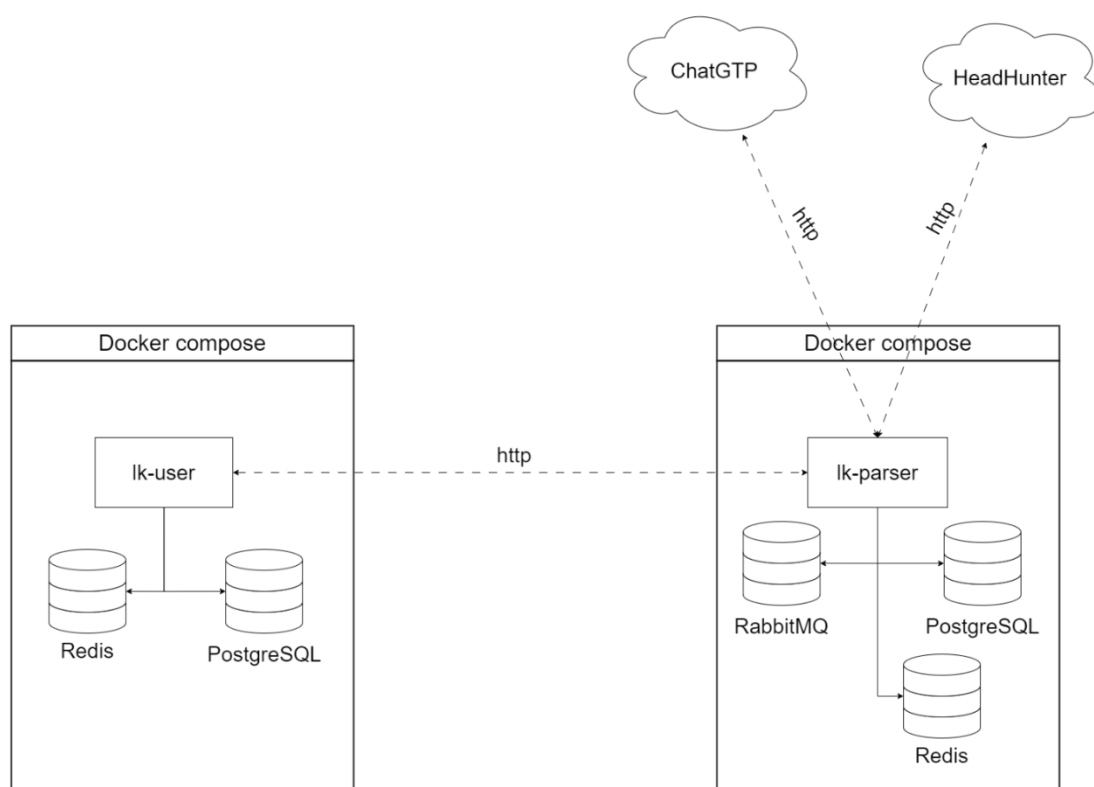


Рисунок 5 – Архитектура взаимодействия сервисов

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

Для хранения полезных данных была предложена следующая схема базы данных сервиса личных кабинетов (рисунок А.1).

База данных сервиса личного кабинета состоит из 6 сущностей:

- а) пользователь – чтобы обеспечивать ограниченный доступ к ресурсу;
- б) миграции – для удобства развертывания приложения и актуализации базы данных при разработке;
- в) hh–регион – регионы хранятся в базе данных чтобы не запрашивать их с hh при каждом формировании аналитического запроса;
- г) hh–индустрия – аналогично региону;
- д) hh–подиндустрия – аналогично региону, связана с hh–индустрией первичным ключом;
- е) аналитический запрос – аналитический запрос хранится на стороне личного кабинета чтобы можно было переотправлять запросы при неудачном выполнении.

На рисунке Б.1 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.

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:

Хотя jQuery многими считается устаревшей технологией – она не теряет своих многочисленных плюсов, которые перечислены в официальной документации [8]:

а) простота использования – jQuery предлагает простой синтаксис, что делает работу с DOM и событиями веб-страницы более интуитивной;

б) кроссбраузерная совместимость обеспечивает однородное поведение на различных браузерах, что делает выполнение кода стабильным;

в) богатая библиотека функций предлагает широкий спектр функций и плагинов для анимации, AJAX запросов, управления CSS и других задач, упрощая разработку;

г) эффективность позволяет создавать компактный и эффективный код за счёт методов выбора элементов DOM и массовой манипуляции элементами;

д) активное сообщество и поддержка – jQuery имеет огромное сообщество разработчиков и обширную документацию, обеспечивая доступ к множеству ресурсов для обучения и развития.

Такой же участи удостоен PHP, как язык, которого принято считать устаревшим, однако нужно просто правильно его использовать, тогда его недостатки становятся его плюсами. Вот, что пишут про этот язык сами его создатели в документации [9]:

а) PHP легко изучить и использовать для создания веб-приложений;

б) поддержка различных баз данных и богатый функционал делают PHP мощным инструментом для разработки;

в) PHP успешно применяется в создании крупных и масштабируемых веб-приложений;

г) большое сообщество и ресурсы PHP обеспечивает доступ к поддержке и развитию.

Тогда стоит приступить к обсуждению взаимодействия связки этих технологий.

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. Следуя официальной документации Webpack [10], интеграция с Webpack позволяет управлять зависимостями JavaScript и оптимизировать их загрузку, а Bootstrap, судя по документации [11], 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.

2.4 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. Данное техническое задание было передано программисту для разработки сервиса аналитики, который был разработан параллельно этой работе. Сама форма технического задания была согласована с заказчиком и соответствует требованиям, принятым в компании. Она является основой для последующей работы над проектом и обеспечивает единое понимание задачи как со стороны разработчика, так и со стороны заказчика.

В таблице 3 перечислены ключевые термины, необходимые для более глубокого понимания содержания технической документации. Эти термины являются основой для освоения материала и обеспечивают понимание основных концепций и терминологии, используемой в документе.

Таблица 3 – Термины и определения технического задания

Термин / Сокращение	Определение / Расшифровка
Отчет	Файл в формате <code>xlsx</code> , 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

В таблице В.1 содержится информация, которая соответствует одной вакансии в итоговом отчете, который представляет собой `xlsx` файл.

В дополнение к таблице В.1 можно добавить, что в отчете содержится информация об:

- а) дополнительной информации по заработной плате;
- б) медиане по выборке;
- в) всех модах по выборке;
- г) средней по выборке;
- д) перцентилях по 25%.

Контекст запроса к лингвистической модели — приведен в виде скриншота из интерфейса лингвистической модели на рисунке 6.

```

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 (функциональность вакансии), keySkills – набор ключевых навыков из вакансии в виде массива, grade – определи 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.

```

Рисунок 6 – запрос к лингвистической модели

В таблице Г.1 представлен запрос, который поступает в сервис аналитики из формы запроса сервиса личных кабинетов

2.5 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 актуальные вакансии по этому названию. Каждая вакансия соответствует строке xlsx файла. Сама строка формируется путем сбора следующей информации из описания вакансии:

- 1) извлечение задач и функционала;
- 2) извлечение информации о дополнительных условиях труда;
- 3) извлечение результирующего уровня заработной платы,
- 4) извлечение дополнительных данных для верстки отчета.

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

Тогда определим слабые стороны старого процесса:

- а) над обработкой ролей работает несколько человек;
- б) каждая отдельная вакансия обрабатывается в несколько сложных этапов;
- в) основное время уходит на рутинные действия, которые легко автоматизировать.

Следовательно, в новом процессе эти недостатки должны быть исправлены. Тогда новый процесс должен выглядеть как на рисунке 7:

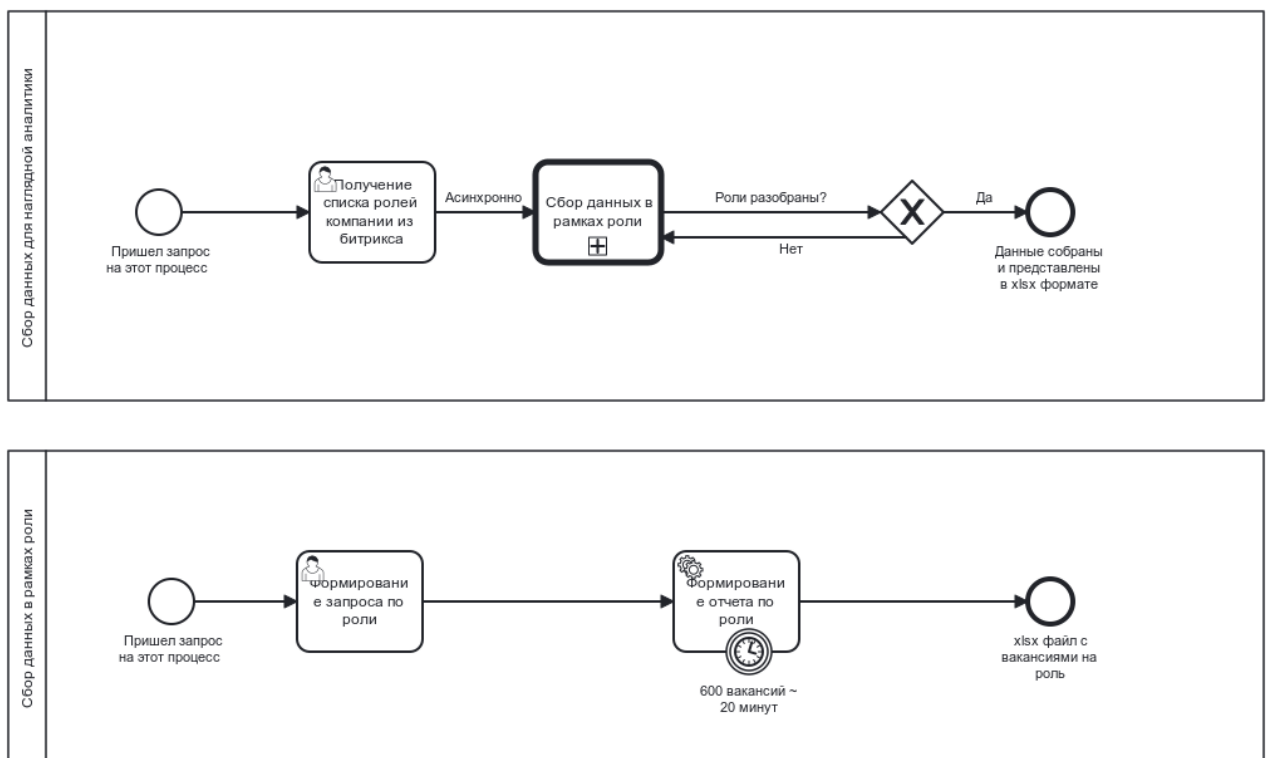


Рисунок 7 – обновленный процесс получения отчета

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.

2.6 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. Более подробно это рассмотрено в одноименной статье [12], судя по ней – вычисления происходят 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, которую OpenAI предоставляет свободно, судя по документации [13].

Т.к. прямые запросы в ChatGPT из России невозможны – нужно использовать прокси для запросов. Таким образом выбор пал на *neuroapi*, где стоимость запроса в 500 токенов (примерно столько занимает запрос одной вакансии) равна 0,02 рубля, исходя из их прейскуранта [14]. То есть при среднем объеме отчета в 600–1000 вакансий (запросов) и 120 должностях. Получаем расход, соответствующий диапазону $600 * 120 * 0.02 \leq x \leq 1000 * 120 * 0.02$ или $1440 \leq x \leq 2400$ рублей за одну реализацию процесса по актуализации заработной платы сотрудников «Artsoft»

Сервер с видеокартой для реализации LLM нейросети (или же упрощенной NLP – Natural Language Processing) стоит в месяц минимум 7500 рублей (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 возможность «экстренного парсинга», также это накладывает лишние расходы на тестирование и разработку продукта.

2.7 User-flow

Для демонстрации работы с продуктом было принято решение воспользоваться руководством по созданию user-flow [15]. User-flow (поток пользователя) представляет собой последовательность действий, которые пользователь выполняет при взаимодействии с продуктом или сервисом (рисунок 8). 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.).

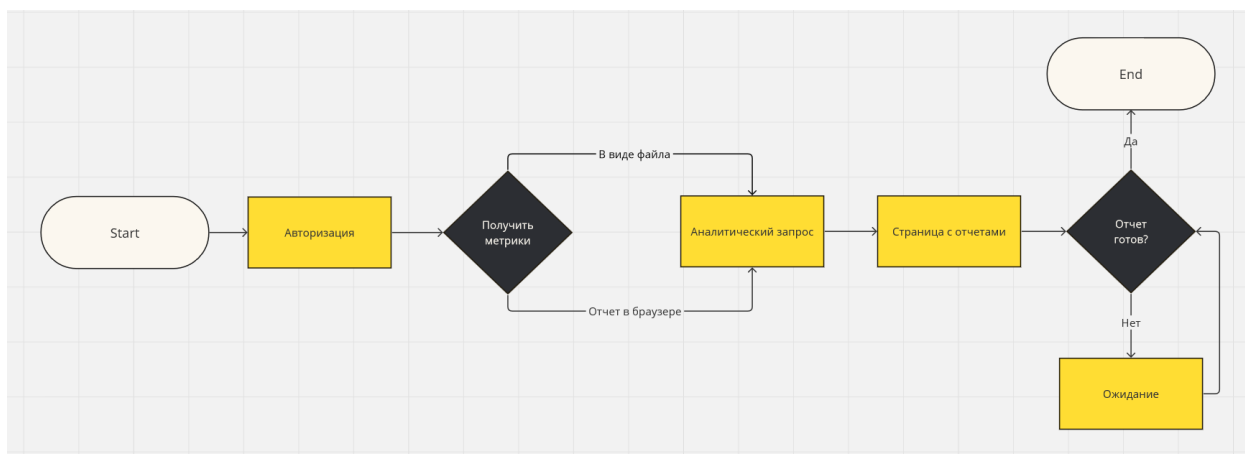


Рисунок 8 – user flow получения отчета

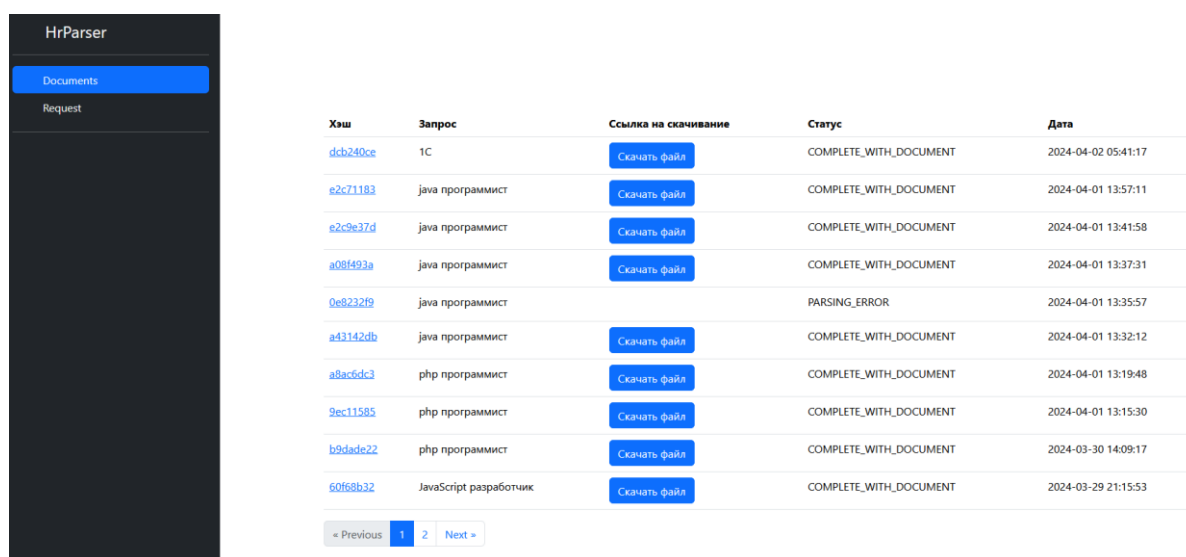
Принцип работы с нашим решением довольно прост и не требует дополнительного обучения персонала: чтобы получить отчет необходимо войти в личный кабинет, используя логин и пароль, далее нужно зайти во вкладку Request (рисунок 9) и заполнить нужные для запроса поля.

Рисунок 9 – форма запроса

Запрос описан в таблице Г.1 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 10) и увидеть статус документа:

- а) START_PARSING – парсинг начат;
- б) COMPLETE – парсинг окончен;
- в) GENERATE_DOCUMENT – генерация документа;
- г) COMPLETE_WITH_DOCUMENT – генерация документа окончена;
- д) PARSING_ERROR – ошибка при выполнении.



Хэш	Запрос	Ссылка на скачивание	Статус	Дата
dcb240ce	1С	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-02 05:41:17
e2c71183	java программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:57:11
e2c9e37d	java программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:41:58
a08f493a	java программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:37:31
0e8232f9	java программист		PARSING_ERROR	2024-04-01 13:35:57
a43142db	java программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:32:12
a8ac6dc3	php программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:19:48
9ec11585	php программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-04-01 13:15:30
b9d9de22	php программист	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-03-30 14:09:17
60f68b32	JavaScript разработчик	Скачать файл	COMPLETE_WITH_DOCUMENT	2024-03-29 21:15:53

[« Previous](#) [1](#) [2](#) [Next »](#)

Рисунок 10 – страница со списком отчетов

Также при нажатии на хэш можно посмотреть запрос более детально (рисунок 11), что позволяет получить дополнительную информацию о конкретном запросе или его результате. Это особенно полезно для анализа данных и выявления паттернов или аномалий в информации.

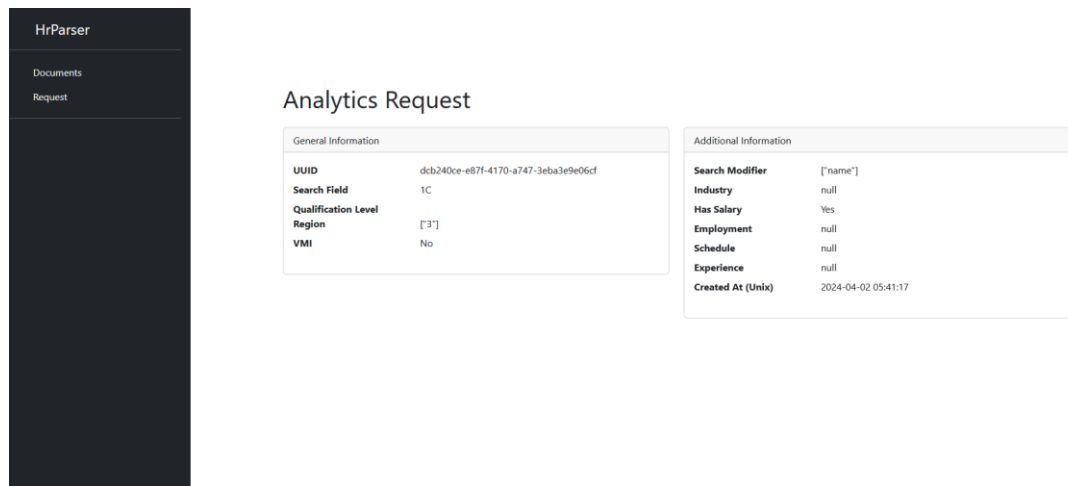


Рисунок 11 – страница с содержанием запроса

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.

При разработке системы были выполнены следующие задачи:

- а) изучение процессов организации встречи в компании Artsofte;
- б) выявление и описание требований к системе у заказчика;
- в) описание функциональных требований к системе;
- г) описание сценариев использования;
- д) подготовка и передача технического задания в разработку.

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. TalentTech: офиц. сайт. – Москва. – URL: <https://main.talenttech.ru/> (дата обращения: 05.04.24)
2. Хабр (М.Тех – М.Видео–Эльдорадо): офиц. сайт. – Москва. – URL <https://habr.com/ru/companies/mvideo/articles/732966/> (дата обращения 05.04.24).
3. Информация по конкурентному сервису Хабр–карьера: офиц. сайт. – Москва. – URL: <https://career.habr.com> (дата обращения 05.04.24).
4. Информация по конкурентному сервису hh.ru: офиц. сайт. – Москва. – URL: <https://hh.ru> (дата обращения 05.04.24).
5. Информация по конкурентному сервису Rit.work: офиц. сайт. – Москва. – URL: <https://rit.work> (дата обращения 05.04.24).
6. Информация по конкурентному сервису Роснавык: офиц. сайт. – Москва. – URL: <https://lk.rosnavyk.ru> (дата обращения 05.04.24).
7. Информация по конкурентному сервису SuperJob: офиц. сайт. – Москва. – URL: <https://www.zarplatomer.ru> (дата обращения 05.04.24).
8. jQuery: офиц. сайт. – Сан-Франциско. – URL: <https://jquery.com/> (дата обращения 05.04.24).
9. PHP: офиц. сайт. – Сан-Франциско. – URL: <https://www.php.net/> (дата обращения 05.04.24).
10. Webpack: офиц. сайт. – Сан-Франциско. – URL: <https://webpack.js.org/> (дата обращения 05.04.24).
11. Build fast, responsive sites with Bootstrap: офиц. сайт. – Кембридж. – URL: <https://getbootstrap.com/> (дата обращения 05.04.24).
12. Локальные нейросети (генерация картинок, локальный chatGPT): офиц. сайт. – Москва. – URL: <https://habr.com/ru/articles/727834/> (дата обращения 05.04.24).
13. ChatGPT openAPI: офиц. сайт. – Сан-Франциско. – URL: <https://chat.openai.com/> (дата обращения 05.04.24).

14. NeuroAPI: офиц. сайт. – Москва. – URL: <https://neuroapi.host/> (дата обращения 05.04.24).

15. User flow: как создаются популярные приложения и сайты: офиц. сайт. – Москва. – URL: <https://habr.com/ru/articles/496760/> (дата обращения 05.04.24).

ПРИЛОЖЕНИЕ А **(обязательное)** **Схема базы данных сервиса ЛК**

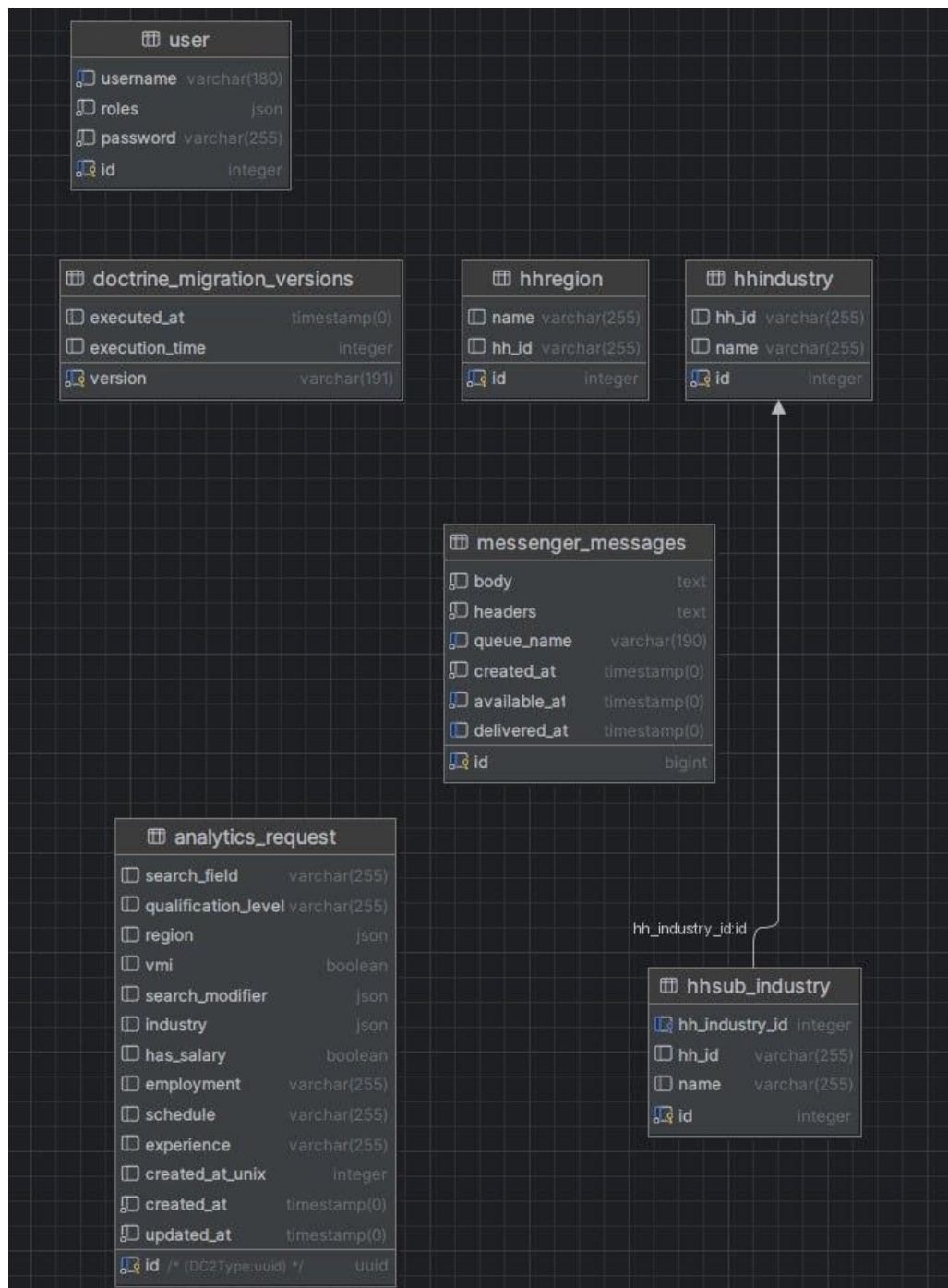


Рисунок А.1 – База данных сервиса личных кабинетов

(обязательное)

Схема базы данных сервиса аналитики

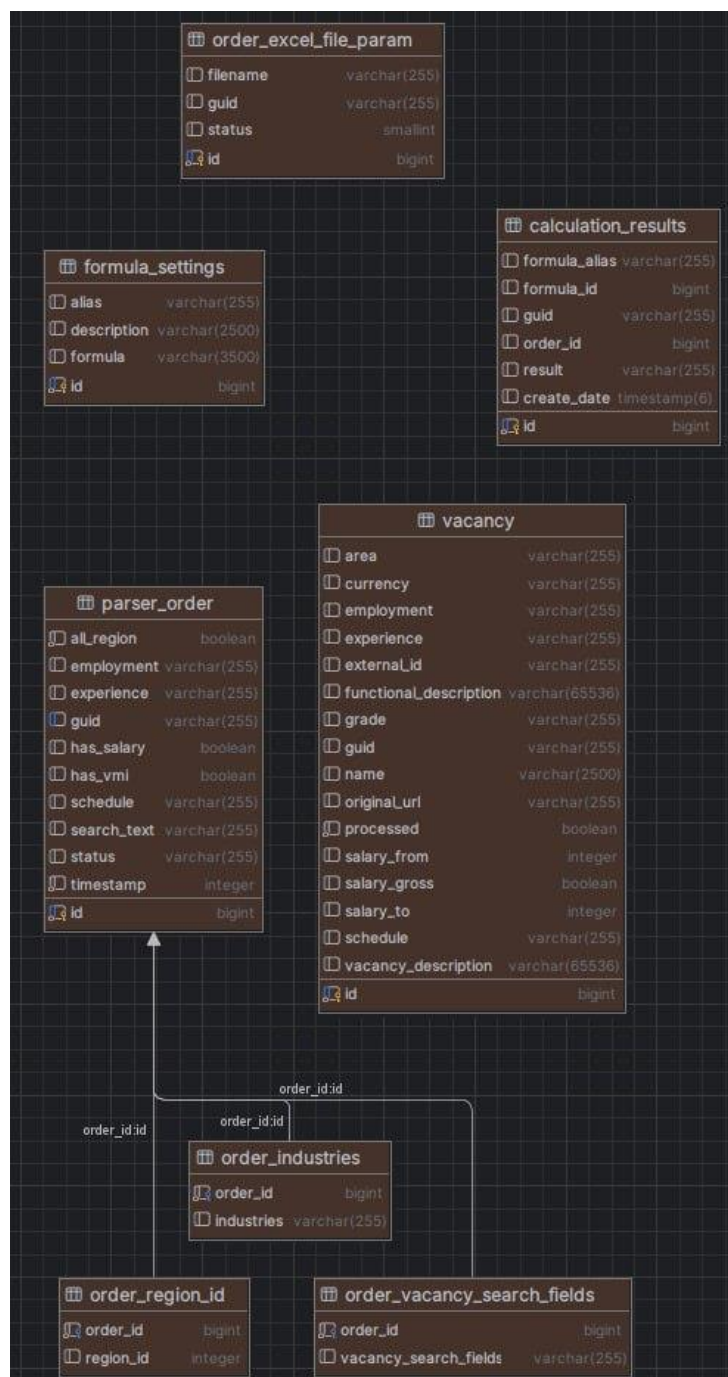


Рисунок Б.1 – База данных сервиса аналитики

ПРИЛОЖЕНИЕ В
(обязательное)
Содержание xlsx файла (отчета)

Таблица В.1 – Содержание xlsx файла (отчета)

Столбец	Пояснение	Определяет ся нейросетью	Пример
1	2	3	4
Ссылка на вакансию в hh.ru	Прямая ссылка на страницу вакансии на hh.ru	Нет	https://ekaterinburg.hh.ru/vacancy/90947747
Набор навыков из словаря hh.ru	Ключевые навыки, которые работодатель ожидает у соискателя, взятые из словаря hh.ru	Да	Python, SQL, JavaScript, аналитическое мышление, коммуникабельность
Название вакансии	Официальное название вакансии, с площадки hh.ru	Нет	Специалист по разработке программного обеспечения
Грейд	Уровень квалификации, указанный в вакансии	Да	Junior

Продолжение ПРИЛОЖЕНИЯ В

Продолжение таблицы В.1

1	2	3	4
Задача и функционал	Описание основных задач и обязанностей, которые предстоит выполнять сотруднику в данной вакансии	Нет	Разработка и поддержка программного обеспечения, участие в проектировании систем, тестирование и оптимизация кода
Описание вакансии	Описание вакансии, предоставленное работодателем	Нет	Мы ищем опытного специалиста с углубленными знаниями в области программирования для участия в проектах по созданию высокотехнологичных решений

Продолжение ПРИЛОЖЕНИЯ В

Продолжение таблицы В.1

Требования в вакансии	Опыт работы	Нет	Опыт работы от 3 лет, знание Python и опыт работы с базами данных, высшее техническое образование
Наличие ДМС	Информация о наличии медицинской страховки	Да	Да
Рабочее время	Информация о наличии рабочем графике	Да	Гибкий график
Удаленная работа	Информация о возможности удаленной работы	Нет	Да
Результиру ющий уровень заработной платы	Диапазон заработной платы, который предоставляется для данной вакансии	Нет	80 000 – 120 000

Продолжение ПРИЛОЖЕНИЯ В

Продолжение таблицы В.1

Название компании	Название компании, предлагающей вакансию	Нет	ArtSofte
ID компании	ID компании с сайта hh.ru	Нет	1234567

ПРИЛОЖЕНИЕ Г
(обязательное)
Аналитический запрос

Таблица Г.1 – содержание аналитического запроса

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

Продолжение ПРИЛОЖЕНИЯ Г

Продолжение таблицы Г.1

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