# ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук Департамент программной инженерии

|  |  |
| --- | --- |
| СОГЛАСОВАНО  Научный руководитель ОП «Программная инженерия», доцент департамента программной инженерии факультета компьютерных наук, кандидат технических наук  А.В.Меликян  « » 2024 г. | УТВЕРЖДАЮ  Академический руководитель ОП  «Программная инженерия», старший преподаватель департамента программной инженерии  Н.А. Павлочев  « » 2024 г. |

# Программа для анализа рынка труда аналитиков данных

|  |  |
| --- | --- |
| *Подп. и дата* |  |
| *Инв. № дубл.* |  |
| *Взам. инв. №* |  |
| *Подп. и дата* |  |
| *Инв. № подл* |  |

**Техническое задание ЛИСТ УТВЕРЖДЕНИЯ**

# RU.17701729.05.10-01 ТЗ 01-1-ЛУ

**Исполнитель:**

студент группы БПИ2310

А.М. Муслин

« » 2024 г.

УТВЕРЖДЁН RU.17701729.05.10-01 ТЗ 01-1-ЛУ

# Программа для анализа рынка труда аналитиков данных

# Техническое задание RU.17701729.05.10-01 ТЗ 01-1

**Листов: 25**

|  |  |
| --- | --- |
| *Подп. и дата* |  |
| *Инв. № дубл.* |  |
| *Взам. инв. №* |  |
| *Подп. и дата* |  |
| *Инв. № подл* |  |

# АННОТАЦИЯ

Техническое задание – основной документ, оговаривающий набор требований и порядок создания программного продукта, в соответствии с которым производится разработка программы, ее тестирование и приемка.

Настоящее Техническое задание на разработку программы для анализа рынка труда аналитиков данных содержит следующие разделы: «Введение», «Основание для разработки», «Назначение разработки»,

«Требования к программе», «Требования к программной документации», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки» и приложения [[9]](#_bookmark50).

В разделе «Введение» указано наименование программы и объекта, в котором программа используется и краткая характеристика области применения программы и объекта, в котором она используется.

В разделе «Основания для разработки» указан документ, на основании которого ведется разработка и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение программного продукта.

Раздел «Требования к программе» содержит основные требования к функциональным характеристикам, к надежности, к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению, а также специальные требования.

Раздел «Требования к программным документам» содержит предварительный состав программной документации и специальные требования к ней.

Раздел «Технико-экономические показатели» содержит ориентировочную экономическую эффективность, предполагаемую годовую потребность, экономические преимущества разработки программы.

Раздел «Стадии и этапы разработки» содержит стадии разработки, этапы и содержание работ. В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

Настоящий документ разработан в соответствии с требованиями:

1. ГОСТ 19.101-77 Виды программ и программных документов[[3]](#_bookmark44).
2. ГОСТ 19.102-77 Стадии разработки[[4]](#_bookmark45).
3. ГОСТ 19.103-77 Обозначения программ и программных документов[[5]](#_bookmark46).
4. ГОСТ 19.104-78 Основные надписи[[6]](#_bookmark47).
5. ГОСТ 19.105-78 Общие требования к программным документам[[7]](#_bookmark48).
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом[[8]](#_bookmark49).
7. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению[[9]](#_bookmark50).

Изменения к данному Техническому заданию оформляются согласно ГОСТ 19.603-78 [[13],](#_bookmark51) ГОСТ 19.604-78 [[14]](#_bookmark46)[.](#_bookmark37)

# СОДЕРЖАНИЕ

1. [ВВЕДЕНИЕ 6](#_ВВЕДЕНИЕ)
   1. [Наименование программы и объекта, в котором программа используется 6](#1.1._Наименование_программы)
   2. [Краткая характеристика области применения программы и объекта, в](#1.2._Краткая_характеристика_области_прим)

[котором она используется 6](#1.2._Краткая_характеристика_области_прим)

1. [ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ 7](#2._ОСНОВАНИЯ_ДЛЯ_РАЗРАБОТКИ)
   1. [Документ(ы), на основании которого(ых) ведётся разработка 7](#2.1._Документ(ы),_на_основании_которого()
   2. [Наименование темы разработки 7](#_bookmark5)
2. [Назначение разработки 8](#_НАЗНАЧЕНИЕ__)
   1. [Функциональное назначение 8](#3.1._Функциональное_назначение)
   2. [Эксплуатационное назначение 8](#3.2._Эксплуатационное_назначение)
3. [ТРЕБОВАНИЯ К ПРОГРАММЕ 10](#4._ТРЕБОВАНИЯ_К_ПРОГРАММЕ)
   1. [Требования к функциональным характеристикам 10](#4.1._Требования_к_функциональным_характе)
      1. [Требования к базе данных 10](#_Требования_к_базе)
      2. [Требования к серверной части 10](#_Требования_к_серверной)
      3. [Требования к клиентской части 11](#_Требования_к_клиентской)
      4. [Требования к скрипту для сбора данных 11](#_Требования_к_скрипту)
   2. [Требования к надёжности 11](#_Требования_к_надежности)
      1. [Требование к обеспечению надежного функционирования программы 11](#_Требования_к_обеспечению)
      2. [Время восстановления после отказа 11](#_Время_восстановления_после)
      3. [Отказы из-за некорректных действий оператора 12](#_Отказы_из-за_некорректных)
   3. [Условия эксплуатации 12](#_Условия_эксплуатации)
   4. [Требования к составу и параметрам технических средств 12](#_Требования_к_составу)
      1. [Требования к серверному оборудованию 12](#_Требования_к_серверному)
      2. [Требования к клиентскому оборудованию 12](#_Требования_к_клиентскому)
   5. [Требования к информационной и программной совместимости 13](#_Требования_к_информационной)
   6. [Требование к маркировке и упаковке 13](#_Требования_к_маркировке)
   7. [Требования к транспортированию и хранению 13](#_Требования_к_транспортированию)
4. [ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ 14](#_ТРЕБОВАНИЯ_К_ПРОГРАММНОЙ)
   1. [Состав программной документации 14](#5.1._Состав_программной_документации)
   2. [Специальные требования к программной документации 14](#5.2._Специальные_требования_к_программно)
5. [ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ 15](#6._ТЕХНИКО-ЭКОНОМИЧЕСКИЕ_ПОКАЗАТЕЛИ)
   1. [Ориентировочная экономическая эффективность 15](#6.1._Ориентировочная_экономическая_эффек)
   2. [Предполагаемая потребность 15](#6.2._Предполагаемая_потребность)
   3. [Экономические преимущества разработки по сравнению с отечественными и](#6.3._Экономические_преимущества_разработ) [зарубежными аналогами 15](#_bookmark30)
6. [СТАДИИ И ЭТАПЫ РАЗРАБОТКИ 20](#7._СТАДИИ_И_ЭТАПЫ_РАЗРАБОТКИ)
   1. [Стадии разработки, этапы и содержание работ 20](#7.1._Стадии_разработки,_этапы_и_содержан)
   2. [Сроки разработки и исполнители 22](#7.2._Сроки_разработки_и_исполнители)
7. [ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ 23](#8._ПОРЯДОК_КОНТРОЛЯ_И_ПРИЁМКИ)
   1. [Виды испытаний 23](#8.1._Виды_испытаний)
   2. [Общие требования к приёмке работы 23](#8.2._Общие_требования_к_приёмке_работы)

[ПРИЛОЖЕНИЕ 1. СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ 24](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ)

[ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ 26](#ЛИСТ_РЕГИСТРАЦИИ_ИЗМЕНЕНИЙ)

# ВВЕДЕНИЕ

# Наименование программы

Наименование программы – «Веб-приложение DataCareerInsight».

Наименование программы на английским языке – «DataCareerInsight web application».

Краткое наименование программы – «DataCareerInsight».

# Краткая характеристика области применения

“Веб-приложение DataCareerInsight ” - сайт, который позволяет визуализировать информацию о вакансиях и резюме специалистов по аналитике данных в России. Основное применение веб-приложения – предоставление агрегированной информации о рынке труда аналитиков данных, включая разные параметры. Важной частью веб-приложения является интерактивная визуализация, которая дает пользователям возможность выбора критериев для анализа с последующим графическим представлением, а также добавление элементов машинного обучения, а именно линейной регрессии, для прогнозирования ожидаемой зарплаты на основе введенных пользователем параметров.

Основная цель программы — предоставить пользователю возможность самостоятельно анализировать текущее состояние рынка труда аналитиков данных в России, выбирая параметры, которые его интересуют, и получая наглядную и детализированную информацию.

# ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ

# Документ(ы), на основании которого(ых) ведётся разработка

Разработка ведется на основании учебного плана подготовки бакалавров по направлению 09.03.04 «Программная инженерия» и утвержденной академическим руководителем программы темы курсового проекта.

# Наименование темы разработки

Наименование темы проекта на русском языке – «Программа для анализа рынка труда аналитиков данных».

Наименование темы проекта на английском языке – «Software for Data Analytics Labor Market Analysis».

Условное обозначение темы разработки – «Веб-приложение “DataCareerInsight”»

# НАЗНАЧЕНИЕ РАЗРАБОТКИ

# Функциональное назначение

«Веб-приложение DataCareerInsight» предоставляет пользователям инструменты для анализа текущего состояния рынка труда специалистов в области аналитики данных в России.

Программа предназначена для сбора, обработки и визуализации данных о вакансиях и резюме аналитиков данных с различных источников

Основные функциональные возможности программы включают:

1. Сбор данных о вакансиях и резюме, фильтрация по ключевым параметрам (зарплата, опыт работы, возраст, компания, условия работы и др.);
2. Интерактивные графики и диаграммы для визуализации статистики по выбранным критериям;
3. Хранение и обработка данных в базе данных с помощью SQL, что позволяет эффективно управлять информацией и предоставлять быстрый доступ к данным для пользователей;
4. Кластеризация данных, позволяющая выделить группы специалистов по различным характеристикам (возраст, зарплата, опыт, образование и др.);
5. Прогнозирование заработной платы с помощью линейной регрессии на основе введенных пользователем параметров;
6. Создание дашбордов для динамического отображения данных и анализа трудового рынка аналитиков в режиме реального времени;

# Эксплуатационное назначение

«Веб-приложение DataCareerInsight» будет использоваться как инструмент для анализа трудового рынка аналитиков данных для различных категорий пользователей:

1. Соискатели могут использовать программу для оценки рыночных условий, прогнозирования своей заработной платы и оценки востребованных навыков в сфере аналитики данных.
2. Работодатели могут анализировать текущее состояние рынка труда, прогнозировать зарплатные ожидания кандидатов, а также оптимизировать требования и условия для вакансий, чтобы привлечь более квалифицированных специалистов
3. Интересующиеся могут использовать программу для понимания рынка труда в сфере аналитики данных.

Веб-приложение будет доступно через интернет и сможет работать на различных устройствах, включая ПК, планшеты и смартфоны, что обеспечивает максимальную доступность для всех категорий пользователей.

# ТРЕБОВАНИЯ К ПРОГРАММЕ

# Требования к функциональным характеристикам

# Программа должна состоять из трех основных компонентов: базы данных, серверной и клиентской части. Между этими компонентами должно быть налажено взаимодействие для обеспечения обработки данных, их анализа и визуализации. Также будет реализован скрипт для сбора информации с hh.ru[[17]](#_bookmark46) и (опциально) других источников.

# Требования к базе данных

# База данных должна быть реляционной

# Должна хранить информацию о вакансиях (название, уровень зарплаты, требуемые навыки и т.д.)

# Должна поддерживать возможность обновления данных при повторном запуске скрипта для сбора информации

# Должна обеспечивать:

# Уникальность записей

# Быстрый доступ к данным для выполнения аналитических запросову

# Будет использоваться реляционная база данных PostgreSQL, так как она обеспечит масштабируемость, поддержку сложных аналитических запросов и надежную работу с большим кол-вом данных

# Требования к серверной части

# Серверная часть будет реализована с использованием Python

# Основные функции:

# Прием запросов от клиентской части

# Взаимодействие с базой данных

# Извлечение и обработка данных из базы данных

# Использование библиотеки pandas для взаимодействия и анализа датасета

# Применение алгоритма кластеризации на основе K-means

# Использование библиотек sklearn (для кластеризации) и yellowbrick (для демонстрации и выбора кол-ва кластеров)

# Применение алгоритмов линейной регрессии для прогнозирования зарплаты

# Формирование ответа для клиентской части

# Отправка ответа клиентской части

# Будут использоваться библиотеки pandas для взаимодействия и анализа датасета, sklearn для кластеризации и yellowbrick для демонстрации и выбора кол-ва кластеров.

# Требования к клиентской части

# Клиентская часть должна быть реализована в виде веб-приложения, доступного через браузер.

# Основные функции:

# Предоставление интерфейса для выбора параметров анализа (опыт работы, требуемые навыки и т.д.)

# Возможность выбора сферы анализа: вакансии или резюме

# Отправка запросов на серверную часть и получение данных

# Визуализация данных в интерактивной форме:

# Гистограммы и графики

# Анализ и визуализация кластеров

# Результаты регрессионного анализа для прогнозирования зарплаты

# Таблицы с результатами анализа

# Обеспечение фильтрации и сортировки данных в реальном времени

# Должна быть поддержка взаимодействия с пользователем:

# Валидность введенных параметров

# Информирование о состоянии запросов (загрузка, ошибка и т.д.)

# Требования к скрипту для сбора данных (веб-скрепер)

# Автоматический сбор данных о вакансиях и резюме с сайта hh.ru и (опционально) с других источников

# Обеспечение записи собранной информации в базу данных

# Реализация повторной попытки сбора данных при ошибках соединения и обработку исключений для устойчивой работы

# Требования к надежности

# Требования к обеспечению надежного (устойчивого) функционирования программы

# Пользователю, работающему с программой через веб-браузер должен быть предоставлен непрерывный доступ к веб-приложению, расположенному по определенному url-адресу. Веб-сервис не должен непредвиденно прерывать свою работу.

# Время восстановления после отказа

# В случае отказа работы серверной части и последующей недоступности веб-приложения, время восстановления не должно превышать одни рабочие сутки.

# Отказы из-за некорректных действий оператора

# После запуска программы на сервере отказ программы вследствие некорректных действий оператора должен быть исключен. В том числе должна быть исключена возможность непреднамеренного выключения программы, не связанного с техническими неполадками сервера.

# Условия эксплуатации

# Специального обслуживания не требуется. Пользователю достаточно умение обращаться с файловой системой компьютера и с интернетом на базовом уровне.

# Требования к составу и параметрам технических средств

# Требования к серверному оборудованию

Рекомендуемые требования к серверному оборудованию для корректной работы приложения при любой нагрузке:

1) 6 виртуальных ядер процессора Intel Xeon E5 2620 или AMD EPYC 7402P или выше

2) 12 ГБ оперативной памяти

3) 32 ГБ SSD хранилища

4) Скорость интернета 9 Гбит/с

Минимальные требования к серверному оборудованию, достаточные для корректной работы приложения при штатной нагрузке

1) 2 виртуальных ядра процессора Intel Xeon E5 2620 или AMD EPYC 7402P

2) 4 ГБ оперативной памяти

3) 16 ГБ SSD хранилища

4) Скорость интернета 3 Гбит/с

# Требования к клиентскому оборудованию

Рекомендуемые требования к клиентскому оборудованию для корректной работы приложения:

1) Процессор не ниже Intel Core i5-4460 или AMD FX-8300

2) 4 ГБ оперативной памяти или больше

3) 2 ГБ видеопамяти или больше

4) Видеокарта не ниже Nvidia GeForce GTX 660 Ti или AMD Radeon RX 460

Минимальные требования к клиентскому оборудованию для работы приложения:

1) Процессор Intel Core i3-2100 или AMD Athlon II X2 250

2) 2 ГБ оперативной памяти

3) Видеокарта Intel HD Graphics 2000 или AMD Radeon RX Vega 3

4) Монитор разрешением 1280x720

Общие требования к клиентскому оборудованию для работы приложения:

1) Мышь или совместное указывающее устройство

2) Клавиатура

3) Один из браузеров

a) Google Chrome версии 9 или выше

b) Microsoft Edge версии 12 или выше

c) Mozilla Firefox версии 11 или выше

d) Opera версии 15 или выше

e) Safari версии 5.1 или выше

f) Internet Explorer версии 11 или выше

4) Свободный доступ в интернет со скоростью не менее 10 Мбит/с

5) Windows 7 или выше

# Требования к информационной и программной совместимости

# Будет использоваться база данных PostgreSQL версии 13 и выше. Для скрипта и серверного приложения будет использоваться язык программирования Python (версии 3.9 и выше), также будут использоваться библиотеки psycopg2 (или другие аналоги) для работы с PostgreSQL, pandas, numpy, sklearn, yellowbrick, matplotlib (или другие аналоги для визуализации) для обработки и анализа данных, допускается и применение других библиотек. Веб-приложение должно быть написано на языке JavaScript в среде Node.JS. Страницы сайта должны быть написаны с использование HTML, CSS и JavaScript.

# Требования к маркировке и упаковке

# Не требуется, так как продукт является веб-приложением.

# Требования к транспортированию и хранению

# Регулярные резервные копии базы данных должны храниться на удаленных северах или облачных платформах. Конфигурационные файлы и исходный код приложения должны быть размещены в системе контроля версий ([GitHub](https://github.com/Temikmus/Course_work))

# ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

# Состав программной документации

1. «Веб-приложение DataCareerInsight». Техническое задание (ГОСТ 19.201-78 [[9]](#_bookmark50))
2. «Веб-приложение DataCareerInsight». Пояснительная записка (ГОСТ 19.404-79[[12]](#_bookmark50))
3. «Веб-приложение DataCareerInsight». Программа и методика испытаний (ГОСТ 19.301-79[[10]](#_bookmark50))
4. «Веб-приложение DataCareerInsight». Текст программы (ГОСТ 19.401-78[[11]](#_bookmark50))
5. «Веб-приложение DataCareerInsight». Проверка пояснительной записки на Антиплагиат

# Специальные требования к программной документации

1. Программная документация должна быть выполнена в соответствии с ГОСТ 19.106-78 и ГОСТами к каждому виду документа
2. Программная документация и сама программа сдаётся в электронном виде в формате pdf в одном архиве формата zip или rar
3. Итоговый вариант пояснительной записки должен быть загружен в LMS «НИУ ВШЭ» в модуль

«ВКР/КР» и проверен системой «Антиплагиат» не позднее, чем за 7 календарных дней до защиты. Лист, подтверждающий загрузку, должен быть сдан в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы

1. Программная документация, отчёт по курсовому проекту, отзыв руководителя курсового проекта, отчёт из системы «Антиплагиат», код или ссылка на репозиторий, а также другие необходимые материалы должны быть загружены в систему LMS «НИУ ВШЭ» в личном кабинете в дисциплине «Курсовая работа» одним архивом не позднее, чем за 3 календарных дня до защиты курсовой работы

# ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

# Ориентировочная экономическая эффективность

# В рамках данного курсового проекта экономическая эффективность не предусмотрена.

# Разрабатываемое веб-приложение позволяет существенно сократить время, затрачиваемое на анализ рынка труда аналитиков данных в России. Благодаря удобной системе фильтров, интерактивным графикам и инструментам прогнозирования, пользователи смогут:

# Быстро находить необходимую информацию о текущих требованиях работодателей и уровнях заработной платы

# Оценивать востребованность навыков и компетенций для принятия карьерных решений

# Ознакомиться с ситуацией на рынке труда аналитиков данных

# Продукт предназначен для широкой аудитории, включая студентов, специалистов, HR-аналитиков, что делает его универсальным инструментом для анализа рынка труда.

# Предполагаемая потребность

На данный момент на рынке нет интерактивных инструментов, позволяющих глубоко анализировать российский рынок труда аналитиков данных. Основные существующие ресурсы предоставляют данные только в формате статичных вакансий и резюме или статей, в которых нет возможности визуализировать интересующую пользователю информацию о рынке труда.

Разрабатываемый продукт:

* Упрощает процесс анализа данных за счет интерактивных визуализаций
* Предоставляет пользователям возможность находить актуальную информацию о рынке труда без необходимости вручную обрабатывать многочисленные вакансии

# Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами

# Большинство сервисов, предназначенных для анализа рынка аналитиков данных, представляют собой статьи с проанализированной ситуацией, статичными графиками и выводами, однако такие сервисы не позволяют пользователю настраивать информацию под индивидуальные запросы.

# Такими сервисами являются habr.com[[1]](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ), newhr.org [[15]](#_bookmark50) ,navigator.hse.ru[[16]](#_bookmark50) и т.д.

# Однако также существует и аналоги, такие как зарубежный сервис public.tableau.com[[2]](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ), позволяющие персонализировать анализ под себя.

# Ниже представлена сравнительная таблица функционала аналогов:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Критерий | habr.com | newhr.org | navigator.hse.ru | public.tableau.com | DataCareerInsight |
| Статичные графики: Наличие графиков без возможности изменения данных |  |  |  |  |  |
| Интерактивные графики: Возможность менять параметры визуализации (например, фильтрация, масштабирование) |  |  |  |  |  |
| Возможность настройки фильтров: По таким параметрам, как регион, зарплата, опыт работы, навыки |  |  |  |  |  |
| Прогнозирование зарплаты: Расчет зарплат на основе введенных параметров (например, возраст, опыт, компетенции) |  |  |  |  |  |
| Кластеризация данных: Разделение данных на группы для выявления тенденций (например, по регионам, навыкам) |  |  |  |  |  |
| Обновляемость данных: Есть или же нет (статичные) |  |  |  |  |  |
| Доступ к данным о навыках: Анализ востребованных навыков в профессии |  |  |  |  |  |
| Анализ вакансий: Предоставление информации о текущих вакансиях (зарплата, требования, условия) |  |  |  |  |  |
| Анализ резюме: Возможность изучать данные о кандидатах (например, образование, ожидаемая зарплата) |  |  |  |  |  |
| Географический анализ: Визуализация распределения вакансий по регионам |  |  |  |  |  |
| Информация о работодателях: Предоставление данных о компаниях, предлагающих вакансии (название, тип компании — российская или иностранная) |  |  |  |  |  |
| Поддержка разных типов графиков: Гистограммы, круговые диаграммы, линейные графики |  |  |  |  |  |
| Сравнительные графики: Возможность сравнивать данные по различным критериям |  |  |  |  |  |
| Интуитивно понятный интерфейс: Простота использования для пользователей без технических навыков |  |  |  |  |  |
| Поддержка мобильных устройств: Адаптивный дизайн для смартфонов и планшетов |  |  |  |  |  |
| Доступность на русском языке: Локализация для пользователей из России |  |  |  |  |  |
| Доступ к данным через экспорт: Экспорт результатов в CSV, Excel или другие форматы |  |  |  |  |  |
| Выводы и комментарии: представленный анализ рынка труда |  |  |  |  |  |

# Примечание:

# - присутствие следующего функционала

# - остутствие следующего функционала

# - частичное присутствие следующего функционала

# 

# Ключевые преимущества веб-сервиса DataCareerInsight:

# Широкий функционал с акцентом на интерактивность: В отличие от большинства статичных сервисов, таких как habr.com[[1]](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ), newhr.org[[15]](#_bookmark50) или navigator.hse.ru[[16]](#_bookmark50), DataCareerInsight предоставляет пользователю возможность самостоятельно выбирать параметры анализа. Интерактивные графики и фильтры позволят адаптировать результаты под индивидуальные запросы.

# Учет резюме: Важной отличительной чертой станет наличие данных не только о вакансиях, но и о резюме аналитиков данных. Это позволит анализировать не только требования работодателей, но и предложения со стороны кандидатов, создавая полную картину рынка труда.

# Кластеризация данных: Выделение групп на основе ключевых параметров, таких как уровень зарплат, навыки или регион. Это позволит выявлять тенденции и создавать сегменты рынка.

# Прогнозирование зарплат: Расчет ожидаемого диапазона зарплат на основе пользовательских данных, чего нет у большинства аналогов, включая Tableau[[2]](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ).

# Актуальность веб-сервиса DataCareerInsight:

# Tableau Public[[2]](#СПИСОК_ИСПОЛЬЗУЕМОЙ_ЛИТЕРАТУРЫ) — единственный конкурент с мощным интерактивным функционалом, но он не предоставляет данных о резюме и не включает инструментов кластеризации и прогнозирования.

# Остальные сервисы значительно уступают по функционалу, так как предлагают только статичные графики и готовые отчеты, без возможности анализа данных в реальном времени.

# Таким образом, веб-сервис DataCareerInsight займет важное место среди существующих инструментов анализа рынка труда, так как он сочетает интерактивность, широкий функционал и ориентированность на локальный рынок.

# СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

# Стадии разработки, этапы и содержание работ

Стадии и этапы разработки были выявлены с учётом ГОСТ 19.102-77 [[2].](#_bookmark45)

|  |  |  |
| --- | --- | --- |
| **Стадии**  **разработки** | **Этапы работ** | **Содержание работ** |
| 1.Техническое задание | Подготовительные работы | * Формулировка целей проекта: создание интерактивного веб-приложения для анализа рынка труда аналитиков данных * Определение функций программы: сбор данных, визуализация, прогнозирование * Сбор исходных данных о рынке труда |
| Разработка и утверждение технического задания | * Описание требований к программе (интерактивные графики, фильтры, прогнозирование зарплат) * Определение структуры данных и методов анализа (кластеризация, регрессия) * Подготовка и утверждение ТЗ с научным руководителем |
| 2.Разработка программы | Разработка  веб-скрепера и создание датасета | * Создание Python-скрипта для автоматического сбора вакансий/резюме с hh.ru и (опционально) с других источников * Обработка данных (удаление дубликатов, преобразование зарплат, выделение навыков) * Интеграция данных в базу PostgreSQL |
| Реализация серверной части | * Разработка API для связи клиентской части с базой данных * Реализация функций для фильтрации данных и   аналитики (кластеризация, прогнозирование)   * Тестирование работы API и интеграции с базой |
| Разработка клиентской части | * Создание интерфейса для сайта (меню, описание и т.д.) * Создание интерфейса для выбора параметров анализа (фильтры по регионам, навыкам и т.д.) * Создание интерфейса для кластеризации данных * Создание интерфейса для линейной регрессии * Визуализация данных с использованием JavaScript библиотек * Обеспечение интерактивности (динамическое обновление графиков) |
| 3. Тестирование программы | Тестирование компонентов | * Проверка работы скрипта: корректность сбора и обработки данных * Тестирование серверной части: стабильность API, точность кластеризации и прогнозов * Проверка клиентской части: корректное отображение данных и графиков |
| 4. Завершение проекта | Оформление документации | * Написание пользовательской инструкции по работе с веб-приложением * Описание структуры базы данных, методов анализа и визуализации * Сбор итоговых результатов тестирования |
| Подготовка к защите | * Создание презентации, демонстрации работы веб-приложения * Проведение финальных исправлений и доработок |

# Сроки разработки и исполнители

|  |  |  |
| --- | --- | --- |
| Этап разработки | Срок выполнения | Исполнитель |
| Подготовительные работы | 12.11.2024 – 19.11.2024 | Муслин Артемий Михайлович |
| Разработка и утверждение ТЗ | 20.11.2024 – 04.12.2024 | Муслин Артемий Михайлович |
| Написание веб-скрепера и датасета | 05.12.2024 – 25.12.2024 | Муслин Артемий Михайлович |
| Реализация серверной части | 26.12.2025 – 31.01.2025 | Муслин Артемий Михайлович |
| Разработка клиентской части | 01.02.2025 – 10.03.2025 | Муслин Артемий Михайлович |
| Тестирование | 11.03.2025 – 21.03.2025 | Муслин Артемий Михайлович |
| Подготовка документации | 22.03.2025 – 31.03.2025 | Муслин Артемий Михайлович |
| Подготовка к защите | 01.04.2025 – 15.04.2025 | Муслин Артемий Михайлович |

# Программный продукт (программа и документация) должен быть завершен не позднее XX.04.2025 – утверждённого приказом декана ФКН НИУ ВШЭ срока защиты курсовой работы.

# ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ

# Виды испытаний

# Контроль качества и испытания программного обеспечения будут проводиться в соответствии с документом «Веб-приложение DataCareerInsight. Программа и методика испытаний», который разработан на основе ГОСТ 19.301-79[[9]](#_bookmark50).

# Основные виды испытаний будут проверять соответствие требований, указанных в пунктах 4.1. и 5.2 настоящего технического задания и пунктом. Также будет проводиться тестирование производительности (проверка времени загрузки данных и работы визуализации при увеличении объема записей в базе данных, оценка устойчивости программы при параллельных запросах к серверу) .

# Общие требования к приёмке работы

# Проверка программного продукта осуществляется в соответствии с документом «Веб-приложение DataCareerInsight. Программа и методика испытаний» и пунктом 5.2 технического задания.

# Защита курсового проекта осуществляется комиссии, состоящей из преподавателей департамента программной инженерии, в утвержденные приказом декана ФКН НИУ ВШЭ сроки.

# ПРИЛОЖЕНИЕ 1

**СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ**

1. Анализ рынка труда аналитиков данных [Электронный ресурс] / Практикум от Яндекса. – Режим доступа: <https://habr.com/ru/companies/yandex_praktikum/articles/818111/> (дата обращения: 03.12.2024).
2. Визуализация данных о рынке труда [Электронный ресурс] / Tableau Public. – Режим доступа: <https://public.tableau.com/app/profile/roman4734/viz/BI_15968874028130/sheet0> (дата обращения: 03.12.2024).
3. ГОСТ 19.101-77: Виды программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.102-77: Стадии разработки. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.103-77: Обозначения программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.104-78: Основные надписи. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.105-78: Общие требования к программным документам. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.106-78: Требования к программным документам, выполненным печатным способом. //

Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

1. ГОСТ 19.201-78: Техническое задание. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.301-79: Программа и методика испытаний. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.401-78: Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.404-79: Пояснительная записка. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.603-78: Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.604-78: Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. Исследование рынка труда аналитиков данных – 2023 [Электронный ресурс] / NewHR. – Режим доступа: <https://newhr.org/data/research-analysts-2023#technologies_and_skills> (дата обращения: 03.12.2024).
8. Навигатор образовательных программ [Электронный ресурс] / Национальный исследовательский университет «Высшая школа экономики». – Режим доступа: <https://navigator.hse.ru/> (дата обращения: 03.12.2024).
9. hh.ru [Электронный ресурс]. – Режим доступа: <https://hh.ru/> (дата обращения: 03.12.2024).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ** | | | | | | | | | |
| Номера листов (страниц) | | | | | Всего листов (страниц  ) в докум. | № документа | Входящий  №сопровод ительного докум. и дата | Подп. | Да- та |
| И  з м. | Изменён ных | Заменён ных | Новых | Аннули рованны х |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |