**Спецификация требований к программному обеспечению**

Информация о документе:  
  
Идентификатор документа: SRS-2023-HH  
Редакция: 1.0  
Дата: 2023-02-06  
Автор: Алексей Боровой

[**1. Введение**](#_gjdgxs) **3**

[Цели проекта](#_30j0zll) 3

[Предполагаемая аудитория и последовательность восприятия](#_1fob9te) 3

[Масштаб проекта](#_3znysh7) 3

[Ссылки на источники](#_2et92p0) 3

[**2. Общее описание**](#_tyjcwt) **4**

[Видение продукта](#_3dy6vkm) 4

[Функциональные требования:](#_wkhkbv95vao1) 4

[Классы и характеристики пользователей](#_2s8eyo1) 4

[Среда функционирования продукта (операционная среда)](#_3rdcrjn) 5

[Рамки, ограничения, правила и стандарты](#_26in1rg) 5

[**3. Функциональность системы**](#_1ksv4uv) **5**

[Функциональный блок 1. Сбор данных](#_44sinio) 5

[Функциональный блок 2. Оперативное (промежуточное) хранение данных](#_2jxsxqh) 8

[Функциональный блок 3. Очистка и предобработка информации для помещения в базу данных](#_z337ya) 9

[Функциональный блок 4. Создание базы данных](#_3j2qqm3) 9

[Функциональный блок 5. Помещение обработанной информации в базу данных](#_1y810tw) 9

[Функциональный блок 6. Создание интерактивного дашборда.](#_4i7ojhp) 9

[Функциональный блок 7. Контейнеризация и запуск приложения.](#_2xcytpi) 9

[Функциональный блок 8. Отладка. Поиск и устранение ошибок и неисправностей в работе приложения.](#_1ci93xb) 9

[**4. Требования к внешним интерфейсам**](#_3whwml4) **9**

[**5. Нефункциональные требования**](#_2bn6wsx) **9**

[**6. Прочее**](#_qsh70q) **10**

# Введение

В этом документе представлена полная спецификация требований к программному обеспечению для проекта, направленного на разработку приложения для сбора, обработки, хранения и визуализации информации о вакансиях по специальностям “Аналитик” (Data Analyst) (все направления), “Инженер данных” (Data engineer), “Дата саентист” (Data Scientist).

## Цели проекта

* Получение исчерпывающей и наглядной информации о распределении вакансий по направлениям аналитики данных (в количественном и процентном выражении).
* Определение для каждого направления аналитика данных ключевых навыков, подсчет количественного выражения упоминаний тех или иных навыков в вакансиях, составление рейтинга таких навыков по частоте упоминания.
* Определение распределения вакансий по: географии, отраслям бизнеса, уровню подготовки специалистов (грейд).

## Предполагаемая аудитория и последовательность восприятия

Предполагаемой аудиторией (конечными пользователями) приложения являются соискатели на вакантные должности по направлениям анализа данных.

## Масштаб проекта

Данный проект позволит конечному пользователю получить актуальную информацию о вакансиях в области аналитики, размещенных на сайте hh.ru.

В данном проекте задействована рабочая группа, состоящая из 5 (пяти) сотрудников: лидера команды (Team Lead), системного аналитика, двух дата инженеров и аналитика данных. Бюджет проекта - 0 руб. Сроки проекта - 3 месяца.

Границы проекта: от сайта-агрегатора вакансий ([HH.ru](http://hh.ru)) до конечного потребителя визуальной информации на интерактивном дашборде.

## Ссылки на источники

Вся информация, а также рабочие файлы размещаются в Git-репозитории: <https://github.com/mgarkunov/hh_ru_analytics>

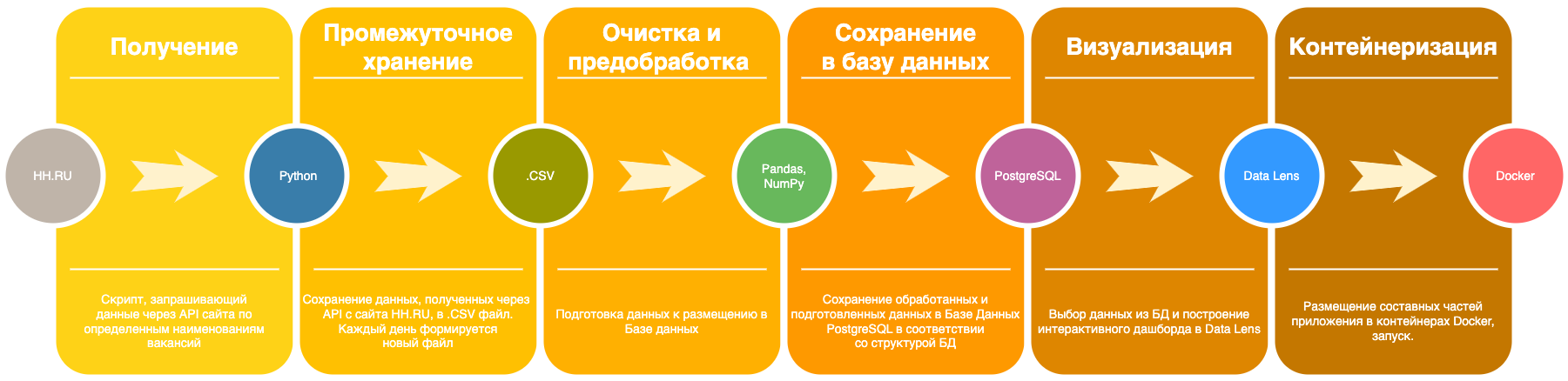
# Общее описание

## Видение продукта

В настоящее время на сайтах-агрегаторах каждый день размещается множество вакансий по различным направлениям бизнеса, промышленности, торговли, услуг и проч. В виду того, что инструментов, которыми должен обладать соискатель на желаемую должность весьма большое количество, такой соискатель может испытывать трудности с определением перечня необходимых к изучению технологий. Данный проект позволит получить распределение по необходимым инструментам и технологиям в зависимости от искомой должности и выделить наиболее популярные у работодателей. Это позволит соискателю чётко определить для себя приоритет в изучении технологий.

Кроме того, в данном проекте проводится аналитика распределения технологий в зависимости от географического расположения работодателя, что позволит сегментировать технологии по населенным пунктам.

В проекте предусматривается систематическая выгрузка данных с сайта-агрегатора, очистка и аккумулирование информации в базе данных, после чего данные размещаются в интерактивном обновляемом дашборде.



## Функциональные требования:

## Классы и характеристики пользователей

Конечный потребитель продукта (пользователь) - соискатель, заинтересованный в получении вакантной должности, либо человек, занимающийся повышением своих навыков, стремящийся получить исчерпывающую информацию по наиболее популярным навыкам, отображаемым в описании вакансии на сайте [HH.ru](http://hh.ru) .

Всего потенциальных пользователей - неограниченное количество.

Предполагается, что основным средством работы с интерактивным дашбордом у пользователей будет являться персональный компьютер.

## Среда функционирования продукта (операционная среда)

Доступ конечного потребителя к интерактивному дашборду должен быть обеспечен из любого веб-браузера и с любого устройства (персональный компьютер, планшет, смартфон), имеющего доступ к сети интернет.

## Рамки, ограничения, правила и стандарты

1. Код, разрабатываемый на языке  Python, должен соответствовать требованиям PEP-8 ([PEP 8 – Style Guide for Python Code](https://peps.python.org/pep-0008/));
2. Запросы к сайту hh.ru при помощи официального API должны выполняться согласно требованиям документации к данному API ([GitHub - hhru/api: HeadHunter API: документация и библиотеки](https://github.com/hhru/api));
3. Проектируемая база данных должна соответствовать требованиям официальной документации разработчика баз данных  PostgreSQL ([15: PostgreSQL 15.1 Documentation](https://www.postgresql.org/docs/current/index.html));
4. Визуализация должна быть выполнена в соответствии с требованиями документации Data Lens ([Yandex DataLens | Yandex Cloud - Документация](https://cloud.yandex.ru/docs/datalens/));
5. Контейнеризация должна производиться согласно требованиям документации  Docker ([Docker Documentation](https://docs.docker.com/));

# Функциональность системы

## Функциональный блок 1. Сбор данных

* Описание и приоритет.

Данный блок представляет собой сбор данных с сайта [HH.ru](https://hh.ru/) через предлагаемый сайтом API с помощью запросов к сайту в среде Python 3.

Приоритет - высокий.

* Функциональные требования:

1. Среда выполнения запроса: Python 3, version 3.8+.
2. Используемые библиотеки: Pandas, NumPy, Requests, Datetime.
3. Скрипт принимает на вход наименование (или список) должности(-ей), согласно утверждённому сайтом [HH.ru](http://hh.ru) алгоритму ([Язык поисковых запросов](https://hh.ru/article/1175)).
4. Сформированы требуемые параметры запроса (ниже указаны только необходимые и достаточные параметры запроса):

* page - номер страницы;
* per\_page - количество элементов на страницу (максимально - 100);
* text - текст фильтра (наименование вакансии или список наименований вакансий);
* experience - наличие опыта (необходимо передавать id из справочника experience);
* employment - тип занятости (необходимо передавать id из справочника employment);
* area - регион поиска (принимается id = 113 (Россия));
* date\_from - дата, которая ограничивает снизу диапазон дат публикации вакансий. Значение указывается в формате ISO 8601 - YYYY-MM-DD;
* date\_to - дата, которая ограничивает сверху диапазон дат публикации вакансий. Значение указывается в формате ISO 8601 - YYYY-MM-DD;
* responses\_count\_enabled - включение счётчика откликов (установить параметр true);

1. Выполнен запрос к сайту через API сайта [HH.ru](http://hh.ru) с помощью библиотеки requests.
2. Получен ответ от сайта (resp.status\_code = 200).
3. Скрипт контролирует количество неуспешных попыток авторизации в рамках одной сессии. В случае превышения лимита попыток авторизации сессия закрывается. Лимит безуспешных попыток авторизации (получения ответа) - 10. Тайм-аут между попытками авторизации 15 секунд.
4. Сформирован датасет с помощью библиотеки pandas. Получены следующие данные по запросу:

| № п/п | Ключ |
| --- | --- |
| 1 | id |
| 2 | premium |
| 3 | billing\_type |
| 4 | relations |
| 5 | name |
| 6 | insider\_interview |
| 7 | response\_letter\_required |
| 8 | area |
| 9 | salary |
| 10 | type |
| 11 | address |
| 12 | allow\_messages |
| 13 | experience |
| 14 | schedule |
| 15 | employment |
| 16 | department |
| 17 | contacts |
| 18 | description |
| 19 | branded\_description |
| 20 | vacancy\_constructor\_template |
| 21 | key\_skills |
| 22 | accept\_handicapped |
| 23 | accept\_kids |
| 24 | archived |
| 25 | response\_url |
| 26 | specializations |
| 27 | professional\_roles |
| 28 | code |
| 29 | hidden |
| 30 | quick\_responses\_allowed |
| 31 | driver\_license\_types |
| 32 | accept\_incomplete\_resumes |
| 33 | employer |
| 34 | published\_at |
| 35 | created\_at |
| 36 | initial\_created\_at |
| 37 | negotiations\_url |
| 38 | suitable\_resumes\_url |
| 39 | apply\_alternate\_url |
| 40 | has\_test |
| 41 | test |
| 42 | alternate\_url |
| 43 | working\_days |
| 44 | working\_time\_intervals |
| 45 | working\_time\_modes |
| 46 | accept\_temporary |
| 47 | languages |

## Функциональный блок 2. Оперативное (промежуточное) хранение данных

* Описание и приоритет.

Данный блок представляет собой алгоритм промежуточного хранения полученных через API данных с сайта [HH.ru](http://hh.ru) .

Приоритет - высокий.

* Функциональные требования:

1. Скрипт, предназначенный для получения данных о вакансиях, после получения всей доступной информации, а также формирования законченного датасета создает файл с расширением .CSV, куда помещает сформированный датасет.
2. Данный .CSV файл получает составное имя, состоящее из:

* Даты, на которую формируется датасет (указывается в параметрах запроса) в формате “YYYY-MM-DD” (генерируется автоматически скриптом Python).
* Наименования “\_temporary\_db.csv”.

Пример: 2023-02-13\_temporary\_db.csv

1. Файл сохраняется на локальном сервере (локальном компьютере), либо подгружается на временное хранение в GitHub (<https://github.com/mgarkunov/hh_ru_analytics>).

## Функциональный блок 3. Очистка и предобработка информации для помещения в базу данных

## Функциональный блок 4. Создание базы данных

## Функциональный блок 5. Помещение обработанной информации в базу данных

## Функциональный блок 6. Создание интерактивного дашборда.

## Функциональный блок 7. Контейнеризация и запуск приложения.

## Функциональный блок 8. Отладка. Поиск и устранение ошибок и неисправностей в работе приложения.

# Требования к внешним интерфейсам

**Интерфейсы пользователя (UX)**

**Программные интерфейсы**

**Интерфейсы оборудования**

**Интерфейсы связи и коммуникации**

# Нефункциональные требования

**Требования к производительности**

**Требования к сохранности (данных)**

**Требования к качеству программного обеспечения**

**Требования к безопасности системы**

**Требования на интеллектуальную собственность**

# Прочее

**Приложение А: Глоссарий**

**Приложение Б: Модели процессов и предметной области и другие диаграммы**

**Приложение В: Список ключевых задач**

| employment | string  Тип занятости. Необходимо передавать id из справочника employment в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). Можно указать несколько значений |
| --- | --- |
| schedule | string  График работы. Необходимо передавать id из справочника schedule в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). Можно указать несколько значений |
| area | string  Регион. Необходимо передавать id из справочника [/areas](https://github.com/hhru/api/blob/master/docs/areas.md). Можно указать несколько значений |
| metro | string  Ветка или станция метро. Необходимо передавать id из справочника [/metro](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-metro-stations). Можно указать несколько значений |
| professional\_role | string  Профессиональная область. Необходимо передавать id из справочника [/professional\_roles](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-professional-roles-dictionary) |
| industry | string  Индустрия компании, разместившей вакансию. Необходимо передавать id из справочника [/industries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-industries). Можно указать несколько значений |
| employer\_id | string  Идентификатор [работодателя](https://api.hh.ru/openapi/redoc#tag/Rabotodatel). Можно указать несколько значений |
| currency | string  Код валюты. Справочник с возможными значениями: currency (ключ code) в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). Имеет смысл указывать только совместно с параметром salary |
| salary | number  Размер заработной платы. Если указано это поле, но не указано currency, то для currency используется значение RUR. При указании значения будут найдены вакансии, в которых вилка зарплаты близка к указанной в запросе. При этом значения пересчитываются по текущим курсам ЦБ РФ. Например, при указании salary=100&currency=EUR будут найдены вакансии, где вилка зарплаты указана в рублях и после пересчёта в Евро близка к 100 EUR. По умолчанию будут также найдены вакансии, в которых вилка зарплаты не указана, чтобы такие вакансии отфильтровать, используйте only\_with\_salary=true |
| label | string  Фильтр по меткам вакансий. Необходимо передавать id из справочника vacancy\_label в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). Можно указать несколько значений |
| only\_with\_salary | boolean  Показывать вакансии только с указанием зарплаты. По умолчанию false |
| period | number  Количество дней, в пределах которых производится поиск по вакансиям |
| date\_from | string  Дата, которая ограничивает снизу диапазон дат публикации вакансий. Нельзя передавать вместе с параметром period. Значение указывается в формате ISO 8601 - YYYY-MM-DD или с точность до секунды YYYY-MM-DDThh:mm:ss±hhmm. Указанное значение будет округлено до ближайших пяти минут |
| date\_to | string  Дата, которая ограничивает сверху диапазон дат публикации вакансий. Нельзя передавать вместе с параметром period. Значение указывается в формате ISO 8601 - YYYY-MM-DD или с точность до секунды YYYY-MM-DDThh:mm:ss±hhmm. Указанное значение будет округлено до ближайших пяти минут |
| top\_lat | number  Верхняя граница широты. При поиске используется значение указанного в вакансии адреса. Принимаемое значение — градусы в виде десятичной дроби. Необходимо передавать одновременно все четыре параметра гео-координат, иначе вернется ошибка |
| bottom\_lat | number  Нижняя граница широты. При поиске используется значение указанного в вакансии адреса. Принимаемое значение — градусы в виде десятичной дроби. Необходимо передавать одновременно все четыре параметра гео-координат, иначе вернется ошибка |
| left\_lng | number  Левая граница долготы. При поиске используется значение указанного в вакансии адреса. Принимаемое значение — градусы в виде десятичной дроби. Необходимо передавать одновременно все четыре параметра гео-координат, иначе вернется ошибка |
| right\_lng | number  Правая граница долготы. При поиске используется значение указанного в вакансии адреса. Принимаемое значение — градусы в виде десятичной дроби. Необходимо передавать одновременно все четыре параметра гео-координат, иначе вернется ошибка |
| order\_by | string  Сортировка списка вакансий. Справочник с возможными значениями: vacancy\_search\_order в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). Если выбрана сортировка по удалённости от гео-точки distance, необходимо также задать её координаты: sort\_point\_lat, sort\_point\_lng. |
| sort\_point\_lat | number  Значение географической широты точки, по расстоянию от которой будут отсортированы вакансии. Необходимо указывать только, если order\_by установлено в distance |
| sort\_point\_lng | number  Значение географической долготы точки, по расстоянию от которой будут отсортированы вакансии. Необходимо указывать только, если order\_by установлено в distance |
| clusters | boolean  Возвращать ли [кластеры для данного поиска](https://github.com/hhru/api/blob/master/docs/clusters.md). По умолчанию — false |
| describe\_arguments | boolean  Возвращать ли [описание использованных параметров поиска](https://github.com/hhru/api/blob/master/docs/vacancies_search_arguments.md). По умолчанию — false |
| no\_magic | boolean  Если значение true — автоматическое преобразование вакансий отключено. По умолчанию – false. При включённом автоматическом преобразовании, будет предпринята попытка изменить текстовый запрос пользователя на набор параметров. Например, запрос text=москва бухгалтер 100500 будет преобразован в text=бухгалтер&only\_with\_salary=true&area=1&salary=100500 |
| premium | boolean  Если значение true — в сортировке вакансий будет учтены премиум вакансии. Такая сортировка используется на сайте. По умолчанию — false |
| responses\_count\_enabled | boolean  Если значение true — дополнительное поле counters с количеством откликов для вакансии включено. По-умолчанию — false |
| part\_time | string  Вакансии для подработки. Возможные значения:   * Все элементы из working\_days в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). * Все элементы из working\_time\_intervals в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). * Все элементы из working\_time\_modes в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). * Элементы part или project из employment в [/dictionaries](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-dictionaries). * Элемент accept\_temporary, показывает вакансии только с временным трудоустройством. Можно указать несколько значений |
| locale | string  Default: "RU"  Example: locale=EN  Идентификатор локали (см. [Локализация](https://api.hh.ru/openapi/redoc#tag/Obshie-spravochniki/operation/get-locales)) |
| host | string  Default: "hh.ru"  Enum: "hh.ru" "rabota.by" "hh1.az" "hh.uz" "hh.kz" "headhunter.ge" "headhunter.kg"  Example: host=hh.ru |