# 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 и обработку полученных данных.

Приоритет - высокий.

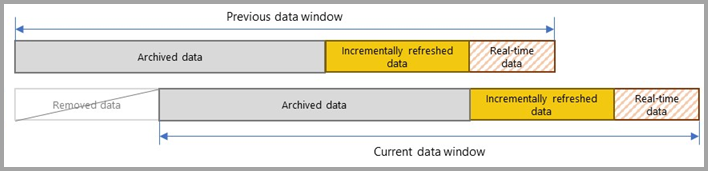
* Функциональные требования:

1. Среда выполнения процедуры: Python 3, version 3.8+. Используемые библиотеки: Pandas, Requests, Datetime.
2. Скрипт должен уметь получать данные о вакансиях с сайта [HH.ru](https://hh.ru/) с помощью его API. Данные должны быть возвращены в формате, который определяется работой API (JSON).
3. Аутентификация API: Скрипт должен иметь возможность аутентифицироваться с API [HH.ru](https://hh.ru/), используя необходимые учетные данные (токен или API-ключи), либо производить запросы и получение данных без аутентификации (в случае, если API это позволяет). Необходимо предусмотреть оба способа.
4. Скрипт должен уметь отправлять запрос на сайт [HH.ru](https://hh.ru/) в соответствии с требованиями документации на  API.
5. Скрипт должен уметь обрабатывать любые ошибки, возникающие во время получения или обработки данных. Это может включать в себя повторное выполнение неудачных запросов или протоколирование сообщений об ошибках для последующего анализа.
6. Скрипт должен уметь обрабатывать получаемые данные в формате JSON, полученные от API [HH.ru](https://hh.ru/), и преобразовывать их в формат плоской таблицы с целью последующего сохранения таких данных в базу данных. Скрипт должен уметь производить определенные операции над данными, такие как: фильтрация, агрегирование или преобразование данных перед их сохранением в базе данных.
7. Скрипт должен быть разработан с учетом производительности, чтобы обеспечить эффективное получение и обработку данных, минимизируя затрачиваемые ресурсы
8. Скрипт должен быть задокументирован, включая информацию о том, как его использовать, любые необходимые зависимости или шаги по настройке, а также ожидаемое поведение и получаемые данные. Это гарантирует, что другие пользователи смогут понять и эффективно использовать скрипт.
9. Запрос на сайт [HH.ru](https://hh.ru/) должен обладать следующими параметрами (ниже указаны только параметры запроса, требующие уточнения, остальные параметры остаются “по-умолчанию”):

* page - номер страницы (значения принимаются равными 1, 2, 3…..n);
* per\_page - количество элементов на страницу (значение принимается равным 100);
* text - текст фильтра (указывается наименование вакансии (только одной));
* area - регион поиска (значение принимается равным 113 (Россия));
* date\_from - дата, которая ограничивает снизу диапазон дат публикации вакансий. Значение указывается в формате ISO 8601: YYYY-MM-DDTHH:MM:SS;
* date\_to - дата, которая ограничивает сверху диапазон дат публикации вакансий. Значение указывается в формате ISO 8601: YYYY-MM-DDTHH:MM:SS;
* responses\_count\_enabled - включение счётчика откликов (параметр принимается true);

ОСТАЛЬНЫЕ ПАРАМЕТРЫ УТОЧНИТЬ (АКТУАЛЬНЫХ ДЛЯ ИЗМЕНЕНИЯ - не найдено)

1. По итогам запроса мы должны выгрузить данные за последние 16 дней и в последующем загружать данные каждый день, периодичностью 1 час. Дабы не было повторов старт выгрузки начинать в 01:00:01, 02:00:01, 03:00:01 и т.д., до 00:00:01. А 00:00:00, 01:00:00, 02:00:0 включать в предыдущий диапазон. Подобный формат используется для того, чтобы уйти от ограничения в 2000 вакансий за одну загрузку.



## Функциональный блок 2. Очистка и предобработка информации для помещения в базу данных

(УЖЕ ВЕРОЯТНО НЕ НУЖЕН, ПОСКОЛЬКУ ДАННЫЕ ОБРАБОТАНЫ СКРИПТОМ В БЛОКЕ 1?)

## Функциональный блок 3. Создание базы данных

Данные могут выгружены в 2-х вариациях: 37 столбцов и 96 столбцов. Основная разниц в том, что в таблице с 37 столбцами «обрезано» выдаются столбцы **requirement** и **responsibility.** Что бы убрать момент неточности информации в данных, будем выгружать вторую таблицу

В базу данных PostgreSQL данные загружать по типам, представленным ниже в таблице:

|  |  |  |  |
| --- | --- | --- | --- |
| **Имя** | **Тип** | **Тип в БД** | **Описание** |
| id | string | INT PRIMARY KEY AUTO\_INCREMENT | Идентификатор вакансии |
| description | string | VARCHAR(300) | Описание вакансии, содержит html |
| branded\_description | string или null | VARCHAR(300) | Брендированное описание вакансии |
| key\_skills | array | VARCHAR(300) | Информация о ключевых навыках, заявленных в вакансии. Список может быть пустым. |
| key\_skills[].name | string | VARCHAR(300) | название ключевого навыка |
| schedule | object | VARCHAR(300) | График работы. Элемент справочника schedule |
| schedule.id | string | INT | Идентификатор графика работы |
| schedule.name | string | VARCHAR(300) | Название графика работы |
| accept\_handicapped | boolean | BOOLEAN(TRUE, FALSE, UNKNOWN) | Указание, что вакансия доступна для соискателей с инвалидностью |
| accept\_kids | boolean | BOOLEAN(TRUE, FALSE, UNKNOWN) | Указание, что вакансия доступна для соискателей от 14 лет |
| experience | object | VARCHAR(300) | Требуемый опыт работы. Элемент справочника experience |
| experience.id | string | INT | Идентификатор требуемого опыта работы |
| experience.name | string | VARCHAR(300) | Название требуемого опыта работы |
| address | object или null | VARCHAR(300) | Адрес вакансии |
| alternate\_url | string | VARCHAR(300) | Ссылка на представление вакансии на сайте |
| apply\_alternate\_url | string | VARCHAR(300) | Ссылка на отклик на вакансию на сайте |
| code | string или null | INT | Внутренний код вакансии работадателя |
| department | object или null | VARCHAR(300) | Департамент, от имени которого размещается вакансия (если данная возможность доступна для компании). Работодатели могут запросить справочник департаментов. |
| department.id | string | INT | Идентификатор департамента |
| department.name | string | VARCHAR(300) | Название департамента |
| employment | object или null | VARCHAR(300) | Тип занятости. Элемент справочника employment. |
| employment.id | string | INT | Идентификатор типа занятости |
| employment.name | string | VARCHAR(300) | Название типа занятости |
| salary | object или null | VARCHAR(300) | Оклад |
| salary.from | number или null | INT | Нижняя граница вилки оклада |
| salary.to | number или null | INT | Верняя граница вилки оклада |
| salary.gross | boolean или null | BOOLEAN(TRUE, FALSE, UNKNOWN) | Признак того что оклад указан до вычета налогов. В случае если не указано - null. |
| salary.currency | string | VARCHAR(50) | Идентификатор валюты оклада (справочник currency). |
| archived | boolean | BOOLEAN(TRUE, FALSE, UNKNOWN) | Находится ли данная вакансия в архиве |
| name | string | VARCHAR(100) | Название вакансии |
| insider\_interview | object или null | VARCHAR(300) | Интервью о жизни в компании |
| area | object | VARCHAR(300) | Регион размещения вакансии |
| area.id | string | INT | Идентификатор региона |
| area.name | string | VARCHAR(300) | Название региона |
| area.url | string | VARCHAR(300) | Url получения информации о регионе |
| initial\_created\_at | string | VARCHAR(300) | Дата и время создания вакансии |
| created\_at | string | VARCHAR(300) | Дата и время публикации вакансии |
| published\_at | string | VARCHAR(300) | Дата и время публикации вакансии |
| employer | object или null | VARCHAR(300) | Короткое представление работодателя. Описание полей смотрите в информации о работодателе. Может не прийти в случае, если вакансия анонимная |
| employer.blacklisted | boolean | BOOLEAN(TRUE, FALSE) | Добавлены ли все вакансии работодателя в список скрытых |
| response\_letter\_required | boolean | BOOLEAN(TRUE, FALSE) | Обязательно ли заполнять сообщение при отклике на вакансию |
| type | object | VARCHAR(300) | Тип вакансии. Элемент справочника vacancy\_type. |
| type.id | string | INT | Идентификатор типа вакансии |
| type.name | string | VARCHAR(300) | Название типа вакансии |
| has\_test | boolean | BOOLEAN(TRUE, FALSE) | Информация о наличии прикрепленного тестового задании к вакансии. В случае присутствия теста - true. |
| response\_url | string или null | VARCHAR(300) | На вакансии с типом direct нельзя откликнуться на сайте hh.ru, у этих вакансий в ключе response\_url выдаётся URL внешнего сайта (чаще всего это сайт работодателя с формой отклика). |
| test | object или null | VARCHAR(300) | Информация о прикрепленном тестовом задании к вакансии. В случае отсутствия теста — null. В данный момент отклик на вакансии с обязательным тестом через API невозможен. |
| test.required | boolean | BOOLEAN(TRUE, FALSE) | Обязательно ли заполнение теста для отклика |
| contacts | object или null | VARCHAR(300) | Контактная информация |
| billing\_type | object | VARCHAR(300) | Биллинговый тип вакансии. Элемент справочника vacancy\_billing\_type. |
| billing\_type.id | string | INT | Идентификатор биллингового типа вакансии |
| billing\_type.name | string | VARCHAR(300) | Название биллингового типа вакансии |
| allow\_messages | boolean | BOOLEAN(TRUE, FALSE) | Включена ли возможность соискателю писать сообщения работодателю, после приглашения/отклика на вакансию |
| premium | boolean | BOOLEAN(TRUE, FALSE) | Является ли данная вакансия премиум-вакансией |
| driver\_license\_types | array | VARCHAR(50) | Список требуемых категорий водительских прав. Список может быть пустым. |
| driver\_license\_types[].id | string | VARCHAR(50) | Категория водительских прав. Элемент справочника driver\_license\_types |
| accept\_incomplete\_resumes | boolean | BOOLEAN(TRUE, FALSE) | Разрешен ли отклик на вакансию неполным резюме |
| working\_days | object или null | VARCHAR(50) | Рабочие дни. Элемент справочника working\_days |
| working\_days.id | string | VARCHAR(50) | Идентификатор рабочих дней |
| working\_days.name | string | VARCHAR(50) | Название рабочих дней |
| working\_time\_intervals | object или null | VARCHAR(50) | Временные интервалы работы. Элемент справочника working\_time\_intervals |
| working\_time\_intervals.id | string | VARCHAR(50) | Идентификатор временного интервала работы |
| working\_time\_intervals.name | string | VARCHAR(50) | Название временного интервала работы |
| working\_time\_modes | object или null | VARCHAR(50) | Режимы времени работы. Элемент справочника working\_time\_modes |
| working\_time\_modes.id | string | VARCHAR(50) | Идентификатор режима времени работы |
| working\_time\_modes.name | string | VARCHAR(50) | Название режима времени работы |
| accept\_temporary | boolean или null | BOOLEAN(TRUE, FALSE, UNKNOWN) | Указание, что вакансия доступна для соискателей с временным трудоустройством |
| professional\_roles | array | VARCHAR(300) | Массив объектов профролей. Список может быть пустым. |
| languages | array или отсутствует | VARCHAR(300) | Список языков. Элементы справочника languages. |
| languages[].id | string | INT | Идентификатор языка. |
| languages[].name | string | VARCHAR(50) | Название языка. |
| languages[].level | object | VARCHAR(50) | Уровень знания языка. Элемент справочника language\_level. |
| languages[].level.id | string | VARCHAR(50) | Идентификатор уровня знания языка. |
| languages[].level.name | string | VARCHAR(50) | Название уровня знания языка. |

## Функциональный блок 4. Помещение обработанной информации в базу данных

## Функциональный блок 5. Создание интерактивного дашборда.

## Функциональный блок 6. 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**