

ё

**Программа для ЭВМ**

**«Атом.Поиск: Базовый релиз»**

**Руководство по развертыванию**

На 19 листах

Дата: 12.03.2024

Версия: 1.0

АННОТАЦИЯ

Настоящее руководство по развертыванию (далее – руководство) предназначено для системных администраторов, производящих установку и сопровождение программы для ЭВМ «Атом.Поиск: Базовый релиз» (далее – ). Руководство содержит сведения о назначении, установке и настройке .

Установка должна выполняться квалифицированным специалистом в области системного администрирования.

СОДЕРЖАНИЕ

[1. Общие сведения 4](#_Toc174366157)

[2. Дистрибутив 7](#_Toc174366158)

[3. Настройка программы 8](#_Toc174366159)

[3.1. Настройка переменных окружения программы 8](#_Toc174366160)

[4. Условия установки 15](#_Toc174366161)

[5. Установка 16](#_Toc174366162)

[Перечень принятых терминов и сокращений 19](#_Toc174366163)

1. Общие сведения
   1. «Атом.Поиск: Базовый релиз» – программа для ЭВМ, предназначенная для поиска информации в различных корпоративных информационных системах по документам различных форматов в рамках единого поискового окна.
   2. Программа состоит из следующих сервисов:

сервис greensearch-frontend обеспечивает взаимодействие пользователя с графическим интерфейсом программы и отображение результатов поиска;

сервис greensearch-proxy обеспечивает получение результатов поискового запроса;

сервис spell-fixer обеспечивает предобработку поисковой фразы (раскладка, опечатки и т.д.);

сервис semantic-service обеспечивает получение эмбеддингов по поисковой фразе;

сервис query-builder обеспечивает построение запроса в поисковый движок на основе параметров пользовательского запроса;

сервис file-loader обеспечивает извлечение текста из файлов;

сервис index-formation обеспечивает предобработку и управление индексируемых документов;

сервис greensearch-admin обеспечивает работу административного интерфейса;

сервис search-history обеспечивает сбор и отображение статистики запросов;

сервис entity-extractor-api обеспечивает извлечение сущностей из файлов;

сервис работы с поисковым движком OpenSearch.

* 1. Список сторонних сервисов, с которыми может взаимодействовать программа:

сервис работы с извлечением текста из изображений, например, Атом.Око: Базовый релиз;

сервис работы с извлечением текста из текстовых файлов - Apache Tika;

сервис работы с объектным хранилищем S3, например, Minio;

сервис брокера сообщений, например, RabbitMQ;

сервис систем управления базами данных - PostgreSQL;

сервис работы с web сервером, например, Nginx.

* 1. На Рис. 1 представлена техническая архитектура .

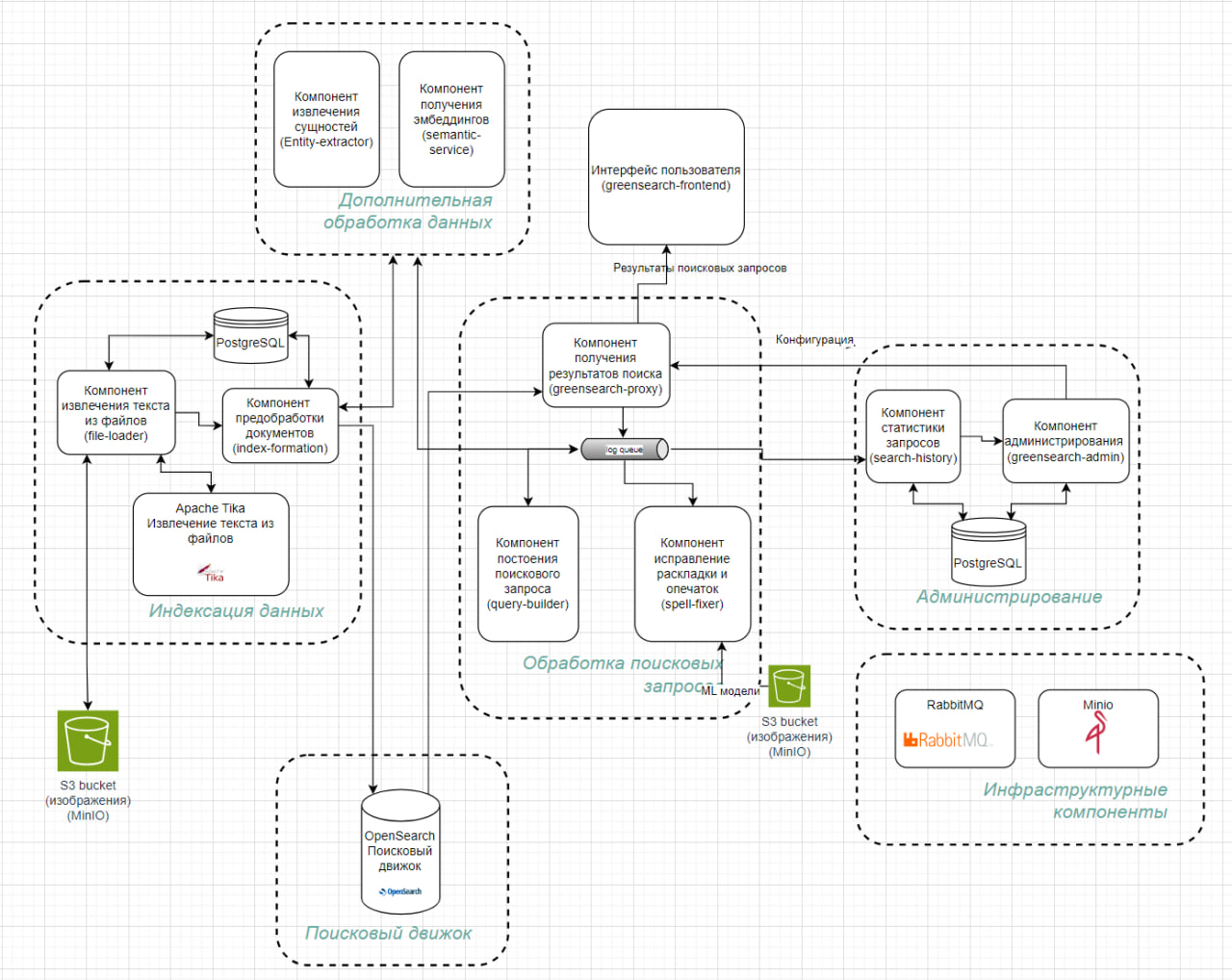


Рис.

1. Дистрибутив
   1. Программа распространяется в виде набора собранных   
      docker-образов, конфигурационных файлов. Образы программы экспортированы в архивы. Состав архивов представлен в таблицах 1 и 2.

Таблица 1

|  |  |
| --- | --- |
| Сервис | Архив |
| greensearch-frontend | greensearch-frontend.tar.gz |
| greensearch-proxy | greensearch-proxy.tar.gz |
| spell-fixer | spell-fixer.tar.gz |
| semantic-service | semantic-service.tar.gz |
| query-builder | query-builder.tar.gz |
| file-loader | file-loader.tar.gz |
| index-formation | index-formation.tar.gz |
| greensearch-admin | greensearch-admin.tar.gz |
| search-history | search-history.tar.gz |
| entity-extractor-api | entity-extractor-api.tar.gz |
| OpenSearch | OpenSearch.tar.gz |

Таблица 2

|  |  |
| --- | --- |
| Настройки дистрибутива | Архив |
| Конфигурационные файлы и дамп БД | config.tar.gz |

1. Настройка программы
   1. Настройка переменных окружения

Описание переменных окружения приведено в таблицах 3 - 13.

* + 1. Сервис greensearch-frontend

Таблица 3

|  |  |  |
| --- | --- | --- |
| Переменная | Описание | Пример |
| BACK\_ADDR\_PROD | Адрес greensearch-proxy | 127.0.0.1:8080 |

* + 1. Сервис greensearch-proxy

Таблица

|  |  |  |
| --- | --- | --- |
| Переменная | Описание | Пример |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| SEARCH\_ENGINE\_HOST | Параметр подключения к сервису OpenSearch - хост | http://search-engine-url.example:9200 |
| SEARCH\_ENGINE\_LOGIN | Параметр подключения к сервису OpenSearch - логин | user |
| SEARCH\_ENGINE\_PASSWORD | Параметр подключения к сервису OpenSearch - пароль | pass |
| AMQP\_URI | Параметр подключения к стороннему сервису работы с брокером сообщений - хост | amqp://rabbitmq:rabbitmq@rmq |

* + 1. Сервис spell-fixer

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| SERVICE\_MODE | Режим работы сервиса: predict - режим сервис исправления запроса; fit - режим обучения ml models | predict/fit |
| MINIO\_LOAD | Загружать модели из/в minio или нет | True/False |
| MINIO\_IP | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 - хост | minio:9000 |
| MINIO\_LOGIN | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -логин | minio-access-key |
| MINIO\_PASS | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -пароль | minio-secret-key |
| MINIO\_INIT\_BUCKET | Minio bucket, в котором хранятся файлы инициализации для моделей | greensearch |
| MINIO\_MODEL\_BUCKET | Minio bucket, в котором будут храниться файлы обученных моделей | greensearch |
| USE\_RPC | Используем ли RPC в сервисе (ставим 0, если не используем gateway и rabbitmq) | True/False |
| AMQP\_URI | Параметр подключения к стороннему сервису работы с брокером сообщений - хост | amqp://rabbitmq:rabbitmq@rmq |
| QUEUE | Очередь для взаимодействия по RabbitRPC | spell-fixer |
| MODELS | Список ml models | MODELS='["inv\_1","inv\_2","jsp\_1"]' |

* + 1. Сервис semantic-service

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| SERVICE\_MODE | Режим работы сервиса: predict - режим сервис исправления запроса; fit - режим обучения ml models | predict/fit |
| MINIO\_LOAD | Загружать модели из/в minio или нет | True/False |
| MINIO\_IP | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -хост | minio:9000 |
| MINIO\_LOGIN | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -логин | minio-access-key |
| MINIO\_PASS | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -пароль | minio-secret-key |
| MINIO\_INIT\_BUCKET | Minio bucket, в котором хранятся файлы инициализации для моделей | greensearch |
| MINIO\_MODEL\_BUCKET | Minio bucket, в котором будут храниться файлы обученных моделей | greensearch |
| USE\_RPC | Используем ли RPC в сервисе (ставим 0, если не используем gateway и rabbitmq) | True/False |
| AMQP\_URI | Параметр подключения к стороннему сервису работы с брокером сообщений - хост | amqp://rabbitmq:rabbitmq@rmq |
| QUEUE | Очередь для взаимодействия по RabbitRPC | spell-fixer |

* + 1. Сервис query-builder

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| AMQP\_URI | Параметр подключения к стороннему сервису работы с брокером сообщений - хост | amqp://rabbitmq:rabbitmq@rmq |
| USE\_RPC | Включен RPC клиент или нет | true/false |
| QUEUE | Наименование очереди для коммуникации по RPC | some-queue |

* + 1. Сервис file-loader

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| MINIO\_ENDPOINT | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 - хост | minio:9000 |
| MINIO\_LOGIN | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -логин | minio\_user |
| MINIO\_PASS | Параметр подключения к стороннему сервису работы с объектным хранилищем S3 -пароль | minio\_password |
| POSTGRES\_URI | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - URL | postgresql+asyncpg://myuser:mypassword@postgres:5432/mydb |
| CELERY\_BROKER\_URL | Параметр подключения к стороннему сервису брокер сообщений - URL брокера для Celery | amqp://rabbitmq:rabbitmq@rmq:5672 |
| OCR\_BROKER\_URL | Параметр подключения для интеграции с OCR - URL брокера для интеграции с OCR | amqp://rmq:rmq@rabbit:5672 |
| OCR\_OUT\_QUEUE | Очередь для исходящих в OCR сообщений | atom\_search\_out |
| OCR\_IN\_QUEUE | Очередь для входящих из OCR сообщений | atom\_search\_in |
| TIKA\_URL | Параметр подключения к стороннему сервису работы с извлечением текста из текстовых файлов - URL | http://tika:9998 |
| INDEX\_FORMATION\_API\_URL | Параметр подключения к сервису index-formation - URL | http://index-formation.example:8000 |

* + 1. Сервис index-formation

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local/dev/prod |
| POSTGRES\_HOST | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - хост | host |
| POSTGRES\_PORT | Порт для подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных | 5432 |
| POSTGRES\_DB | База данных | db\_name |
| POSTGRES\_USER | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - логин | user |
| POSTGRES\_PASSWORD | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - пароль | pass |
| POSTGRES\_SCHEMA | Схема БД | search\_history |
| SEARCH\_ENGINE\_HOST | Параметр подключения к сервису OpenSearch - хост | http://search-engine-url.example:9200 |
| SEARCH\_ENGINE\_LOGIN | Параметр подключения к сервису OpenSearch - логин | user |
| SEARCH\_ENGINE\_PASSWORD | Параметр подключения к сервису OpenSearch - пароль | pass |
| SEARCH\_EMBEDDING\_URL | Параметр подключения к сервису semantic-service - URL сервиса для получения эмбеддингов | http://semantic-service.example |
| CELERY\_BROKER\_URL | Параметр подключения к стороннему сервису брокер сообщений - URL брокера для Celery | <amqp://rabbitmq:rabbitmq@rmq:5672> |

* + 1. Сервис greensearch-admin

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| SECRET\_KEY | Секретный ключ используемый Django для шифрования чувствительной информации в БД | secret |
| ALLOWED\_HOSTS | Список хостов, от которых сервис может принимать запросы (по умолчанию все) | ["0.0.0.0"] |
| DEBUG | Режим отладки | True/False |
| ENVIRONMENT | Наименование окружения, используется для внутренних настроек и для разделения стендов в apm | local/dev/prod |
| SITE\_URL | Host графический интерфейс программы | http://atom-search.example:2024 |
| POSTGRES\_HOST | Параметр подключения к стороннему сервису сервис работы с реляционной подсистемой хранения и извлечения данных - хост | Host СУБД |
| POSTGRES\_PORT | Порт для подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных | 5432 |
| POSTGRES\_DB | База данных | greensearch\_admin |
| POSTGRES\_USER | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - логин | user |
| POSTGRES\_PASSWORD | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - пароль | pass |
| SOURCE\_SEARCH\_HISTORY\_DB\_URL | Ссылка подключения к БД search-history | postgresql+psycopg2://user:pass@db-host/db\_name |

* + 1. Сервис Search-history

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| POSTGRES\_PASSWORD | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - пароль | postgres |
| POSTGRES\_USER | Параметр подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных - логин | postgres |
| POSTGRES\_HOST | Параметр подключения к стороннему сервису сервис работы с реляционной подсистемой хранения и извлечения данных - хост | host.docker.internal |
| POSTGRES\_PORT | Порт для подключения к стороннему сервису работы с реляционной подсистемой хранения и извлечения данных | 5432 |
| POSTGRES\_DB | База данных | search\_history |
| POSTGRES\_SCHEMA | Схема БД | search\_history |
| RABBITMQ\_HOST | Параметр подключения к стороннему сервису брокера сообщений - хост | host.docker.internal |
| RABBITMQ\_PORT | Порт для подключения к стороннему сервису брокера сообщений | 5672 |
| RABBITMQ\_USER | Параметр подключения к стороннему сервису брокера сообщений - логин | rabbitmq |
| RABBITMQ\_PASS | Параметр подключения к стороннему брокера сообщений -пароль | rabbitmq |
| METRICS\_QUEUE | Очередь для сбора истории запросов | search\_query\_metrics |
| REQUEST\_CLICK\_QUEUE | Очередь для сбора кликов пользователей | history |

* + 1. Сервис entity-extractor-api

Таблица

| Переменная | Описание | Пример |
| --- | --- | --- |
| ENVIRONMENT | Наименование окружения | local |

* + 1. Сервис OpenSearch

Таблица 13

| Переменная | Описание | Пример |
| --- | --- | --- |
| OPENSEARCH\_INITIAL\_ADMIN\_PASSWORD | Параметр подключения к сервису OpenSearch - пароль | pass |
| IMAGE\_TAG | Тэг имеджа в docker registry |  |

1. Условия установки

Для серверной части рекомендуется использовать операционную систему Astra Linux Special Edition, также серверная часть может исполняться на любой Linux операционной системе (далее - ОС), поддерживающую технологию контейнеризации.

Тестирование программы выполнялось со следующими версиями программного обеспечения:

Astra Linux, версия 1.6 (Смоленск);

Docker, версия 20.10.14;

Docker-compose, версия 1.29.2.

1. Установка
   1. Установить следующее программное обеспечение:

docker, по инструкции к используемой ОС;

docker-compose, по инструкции к используемой ОС.

* 1. Все примеры установки программы приведены для docker-compose.
     1. Импорт образов программы

Для импорта образов из архивов таблицы 1 необходимо выполнить следующие команды:

$ zcat greensearch-frontend.tar.gz | docker load

$ zcat greensearch-proxy.tar.gz | docker load

$ zcat spell-fixer.tar.gz | docker load

$ zcat semantic-service.tar.gz | docker load

$ zcat query-builder.tar.gz | docker load

$ zcat file-loader.tar.gz | docker load

$ zcat index-formation.tar.gz | docker load

$ zcat greensearch-admin.tar.gz | docker load

$ zcat search-history.tar.gz | docker load

$ zcat entity-extractor-api.tar.gz | docker load

$ zcat OpenSearch.tar.gz | docker load

* + 1. Распаковка архива из таблицы 2

Для распаковки архива с конфигурационными файлами необходимо выполнить следующую команду:

$ tar -xzf ./ config.tar.gz

* + 1. Работа со словарями

Открыть файл ./greensearch/stopwords/stopwords.txt для редактирования и изменить список стоп-слов.

Открыть файл ./greensearch/synonyms/synonyms.txt для редактирования и изменить список синонимов.

* + 1. Запуск docker-compose

Для запуска сервисов программы через docker-compose необходимы команды:

$ cd ./greensearch

$ docker-compose up -d

Docker-compose осуществляет запуск всех сервисов.

* 1. После успешного запуска программа доступна по протоколу HTTP по адресу 127.0.0.1:80 на этом сервере.
  2. Запустить браузер, например, Яндекс Браузер, нажав иконку на рабочем столе (Рис. 2).



Рис.

* 1. Ввести в адресной строке адрес программы - 127.0.0.1.
  2. Открывается окно входа в программу (Рис. 3).

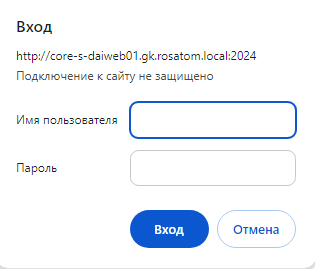


Рис.

* 1. В окне входа в программу необходимо в полях «Имя пользователя» и «Пароль» ввести учетные данные пользователя. Открывается главная страница «Атом.Поиск» (Рис. 4).

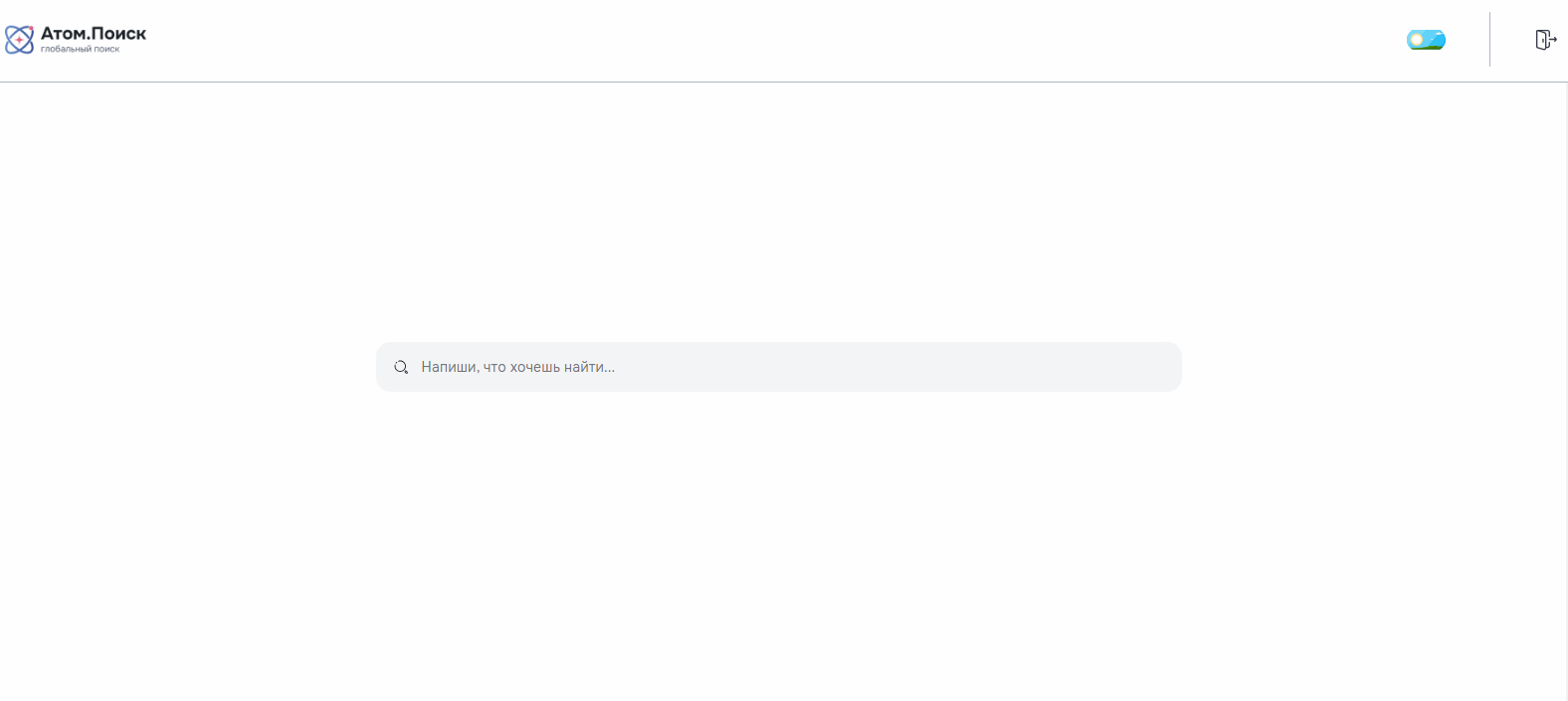


Рис.

Перечень принятых терминов и сокращений

АРМ – Автоматизированное рабочее место

БД – База данных

ОС – Операционная система

Программа – Программа для ЭВМ «Атом.Поиск: Базовый релиз»

СУБД – Система управления базами данных