Пользовательская документация для веб-сервиса «Дайджест» (V 1.2)

Оглавление

Назначение сервиса 2
1. Начало работы с сервисом 5
   1.1. Начало работы пользователя без регистрации 6
   1.2. Регистрация пользователя 6
   1.3. Вход зарегистрированного пользователя 10
2. Основное окно сервиса 12
3. Задачи 13
   3.1. Создание задачи на сбор первичных данных с сайта 14
   3.2. Создание задачи на сбор первичных данных из файла 18
   3.3. Создание задачи на сбор первичных данных из канала Telegram 19
   3.4. Создание задачи на сбор первичных данных из файлов, находящихся на ftp сервере 21
   3.5. Создание задачи на получение дайджеста 22
4. Факты 24
5. Ленты 25
6. Поиск 29
7. Промпты 31
8. Поддержка 31

Назначение сервиса

ЦЕЛЬ: экономия времени пользователя, что предоставляет ему конкурентные преимущества для принятия обоснованных решений.

Веб-сервис «Дайджест» - многофункциональное программное решение в сфере искусственного интеллекта, ориентированное на автоматизацию обработки данных, включая их сбор, интеграцию, анализ, хранение и визуализацию. Веб-сервис «Дайджест» (далее – Платформа) превращает неструктурированный информационный поток в структурированные данные и знания. Платформа предназначена для объединения информации из внутренних и внешних источников, включая сети Интернет, социальные сети и внутренние сети предприятия.

Основные функции сервиса

Сбор данных:

Сервис оснащен инструментами для автоматизированного сбора данных из:

- Внутренних баз данных, включая корпоративные системы управления и хранилища данных;
- Автоматический сбор новостей и статей из большого количества источников: сайты в сети Интернет, блоги, официальные пресс-релизы, Telegram-каналы, FTP-серверы, пользовательские файлы в форматах pdf, doc, docx, txt, xlsx, rtf, jpg, png. Функционал сбора данных включает возможность настройки расписаний для автоматического извлечения информации, использование механизмов парсинга, а также поддержку стандартизированных протоколов (HTTP, FTP).

Интеграция данных:

Интеграция данных обеспечивает их консолидацию и трансформацию для дальнейшего анализа. Сервис поддерживает:

- Преобразование данных в единую структуру с учетом различных форматов и источников;
- Очистку данных от дубликатов, ошибок, нерелевантной информации, рекламы;
- Персонализация ленты: Создание индивидуальной тематической подборки для пользователя на основе его интересов, истории чтения и предпочтений.
- Реферирование: Создание краткого содержания длинных статей, докладов или сводки новостей за период. Может быть экстрактивным (выделение ключевых предложений) или абстрактным (пересказ своими словами).
- Выявление именованных сущностей (Концептов): Автоматическое распознавание и выделение в тексте имен людей, компаний, организаций, брендов, локаций, дат и так далее.
- Инструменты интеграции позволяют объединять данные из множества источников в рамках единой системы, создавая единое информационное пространство.

Хранение данных:

Для хранения данных используются надежные и производительные базы данных. Особенности хранения включают:

- Поддержку реляционных и нереляционных структур данных;
- Репликацию и резервное копирование для обеспечения высокой доступности;

- Возможность работы с большими объемами данных, используя технологии распределенных вычислений.

Преимущества и уникальные особенности

- Модульность и гибкость: возможность настройки и доработки функционала под задачи конкретного клиента.
- Интеграция с внешними источниками: парсинг веб-страниц, обработка данных из социальных сетей, использование внутренних локальных материалов.
- Поддержка больших данных: эффективная работа с объемами информации в десятки терабайт.
- Высокая производительность: оптимизация алгоритмов обработки и анализа данных для минимизации времени выполнения операций.
- Интерактивный интерфейс: удобные инструменты визуализации и взаимодействия для пользователей с разным уровнем технической подготовки.

Примеры применения

Сервис «Дайджест» может использоваться в различных отраслях:

- Бизнес-аналитика: анализ продаж, прогнозирование спроса, выявление трендов.
- Государственное управление: мониторинг социальных процессов, анализ данных о населении.
- Маркетинг: анализ клиентского поведения, управление рекламными кампаниями.
- Научные исследования: обработка данных из научных публикаций, проведение экспериментов.

Расширенные примеры использования:

- Финансовый сектор: анализ транзакционной активности для выявления аномалий и предотвращения мошенничества.
- Здравоохранение: анализ медицинских данных для выявления эпидемиологических тенденций.
- Образование: мониторинг и прогнозирование показателей успеваемости студентов.

1. Начало работы с сервисом

Работа с сервисом доступна как для зарегистрированных пользователей, так и для пользователей без регистрации. Полный функционал сервиса доступен пользователям после регистрации. Для начала работы с сервисом необходимо открыть сайт по адресу https://pressreview.pro. Работа с веб-сервисом начинается с аутентификации пользователя (Рис. 1).

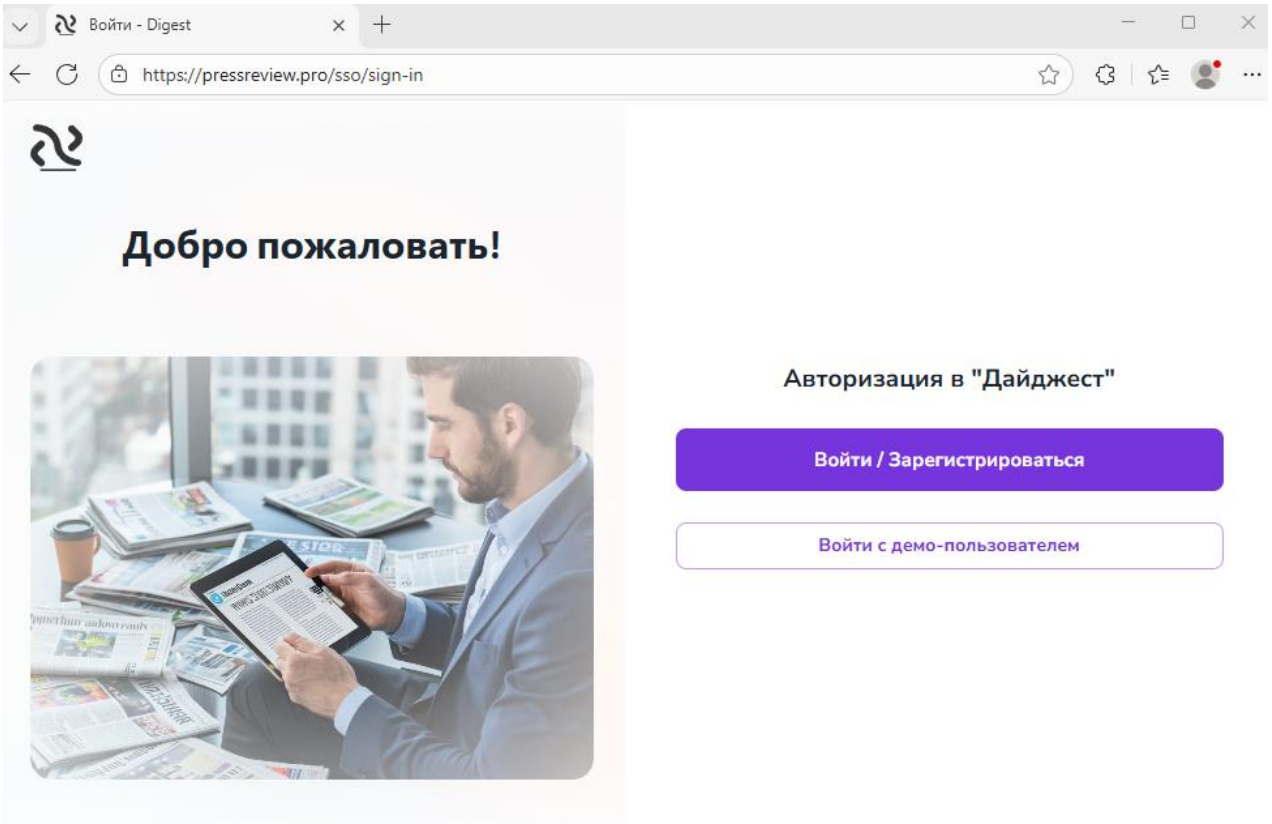


Рис. 1. Стартовая страница веб-сервиса

1.1. Начало работы пользователя без регистрации

Для начала работы пользователя без регистрации необходимо на стартовой странице выбрать пункт меню «Войти с демо-пользователем» (Рис. 2).

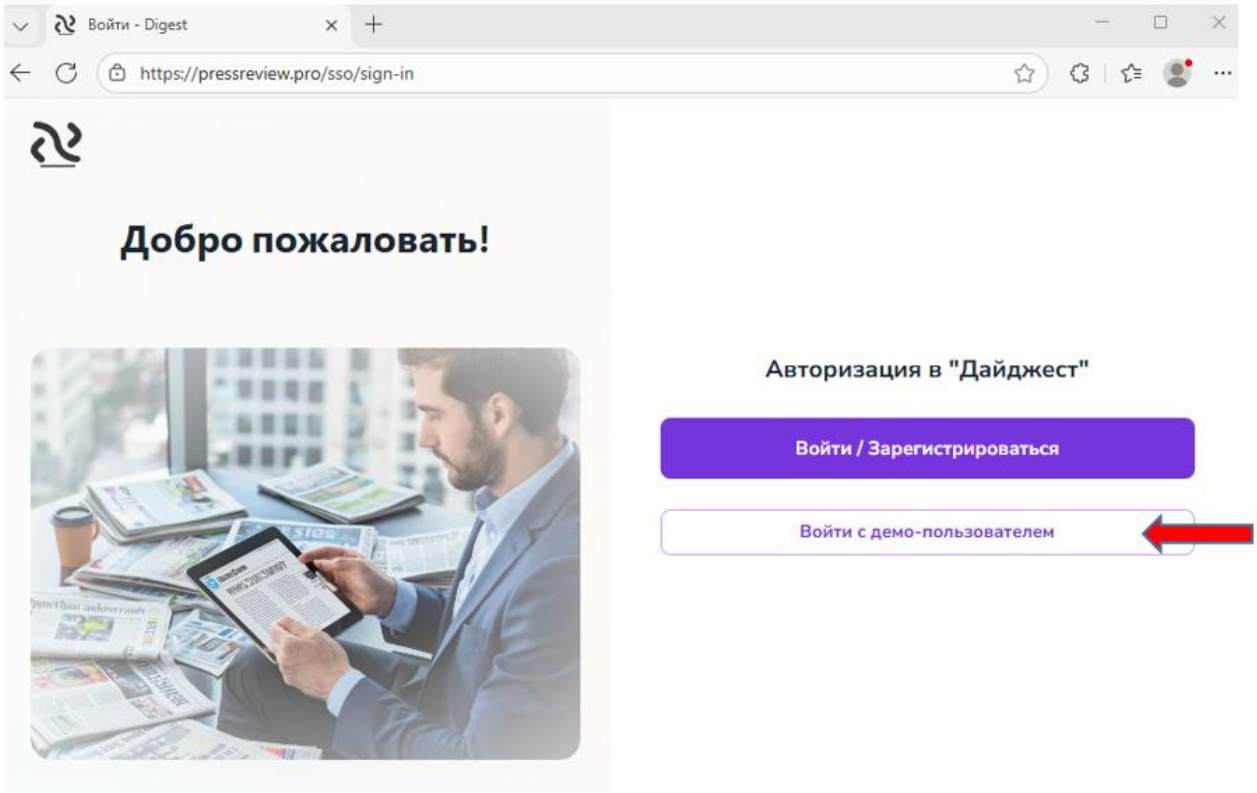


Рис. 2. Вход пользователя без регистрации

После входа доступен демонстрационный функционал сервиса.

1.2. Регистрация пользователя

Для доступа к полному функционалу сервиса необходима регистрация пользователя. (Рис. 3, Рис. 5). Откройте страницу регистрации (Рис. 5).

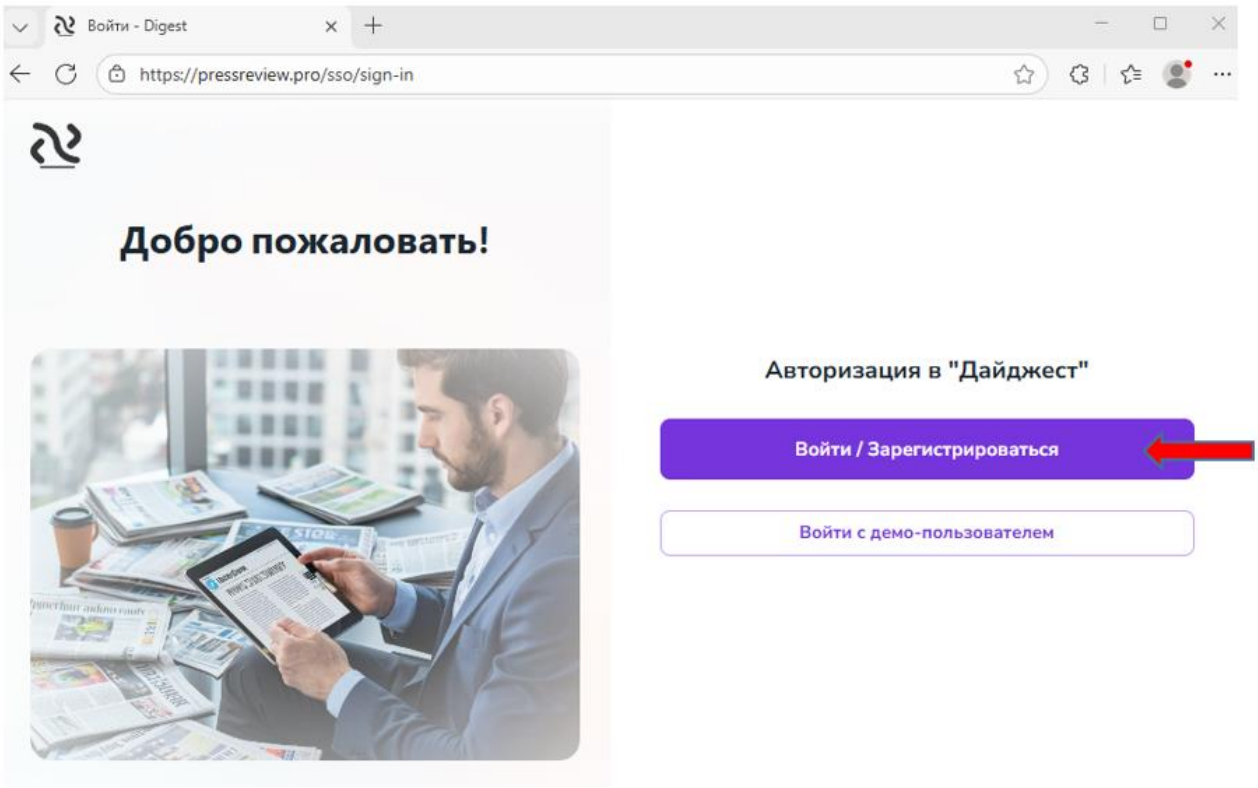


Рис. 3. Вход зарегистрированного пользователя или регистрация

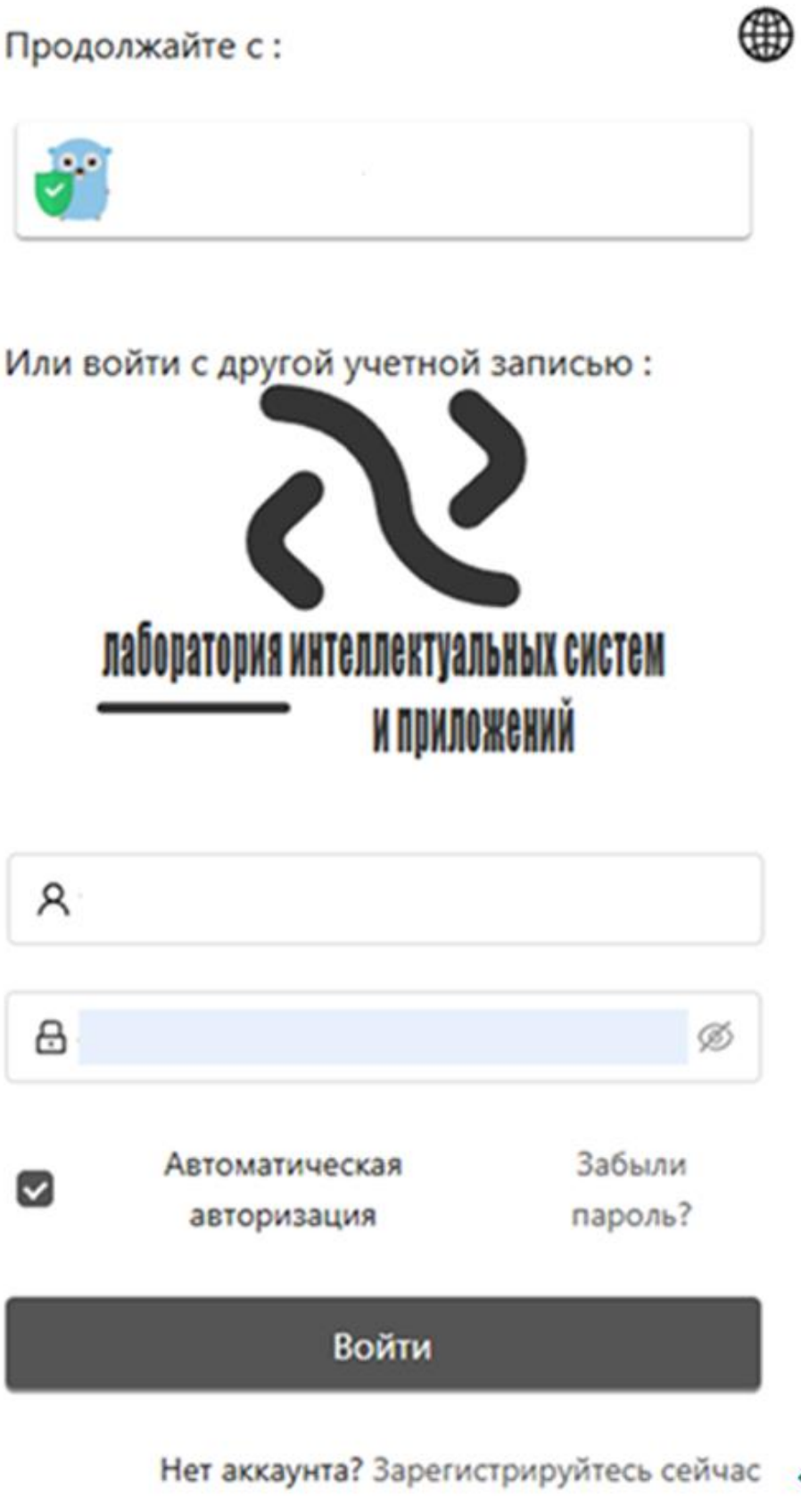

Продолжайте с :
Или войти с другой учетной записью :
лаборатория интеллектуальных систем
и приложений
Автоматическая
авторизация
Забыли
пароль?
Войти
Нет аккаунта? Зарегистрируйтесь сейчас

Рис. 4. Регистрация пользователя в сервисе

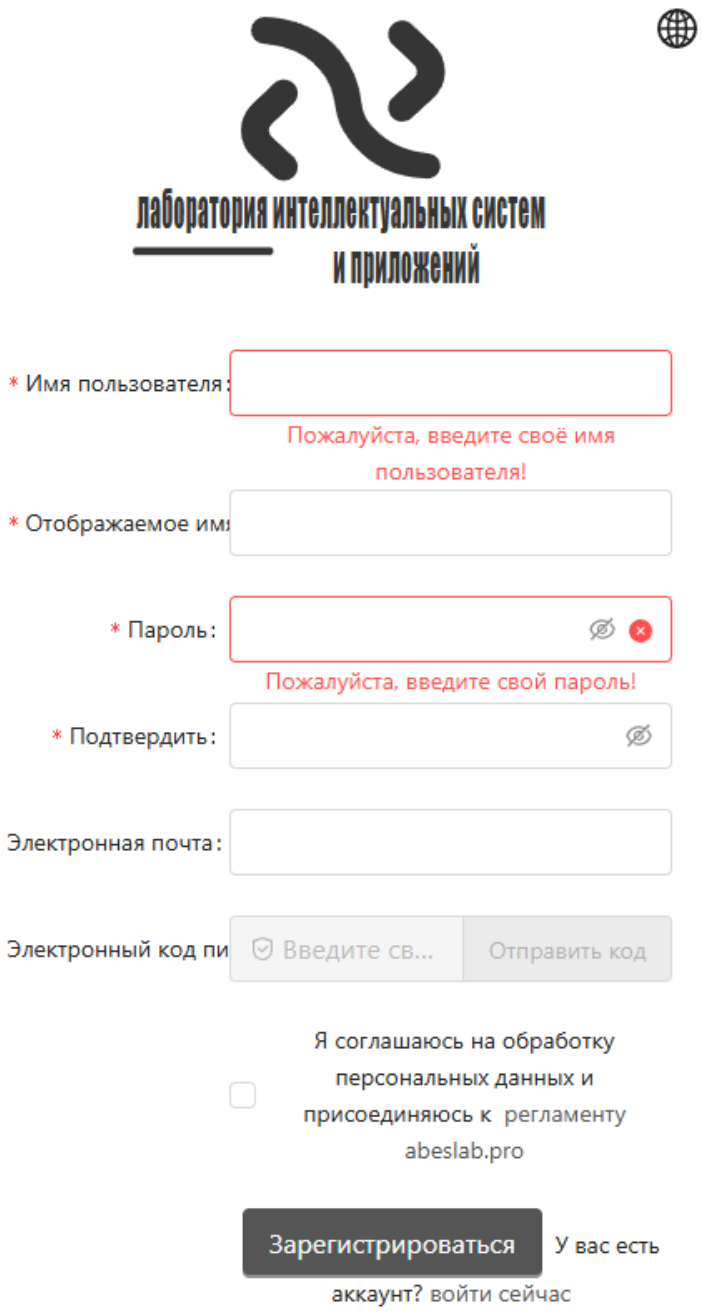


Рис. 5. Страница регистрации

Вам необходимо последовательно заполнить все поля формы:

- Имя пользователя: Введите уникальное имя пользователя (логин).
- Отображаемое имя: Введите имя, которое отображается в системе.
- Пароль: Придумайте и введите надёжный пароль.
- Подтвердить: Повторите введённый выше пароль для проверки.
- Электронная почта: Введите действующий адрес электронной почты.
- Нажмите кнопку «Отправить код».
- Проверьте свою почту и введите полученный цифровой код в поле «Введите свой код».
- Ознакомьтесь с информацией о регламенте работы сервиса и политики обработки персональных данных. Перед завершением регистрации вы должны согласиться с условиями. Поставьте галочку в соответствующем поле.
- Важно: Убедитесь, что все поля заполнены корректно, пароли совпадают, а код с электронной почты введён верно.
- Завершите регистрацию. Нажмите кнопку «Зарегистрироваться».

1.3. Вход зарегистрированного пользователя

Если вы уже зарегистрированный пользователь, то для начала работы необходимо ввести регистрационные данные (Рис. 6).

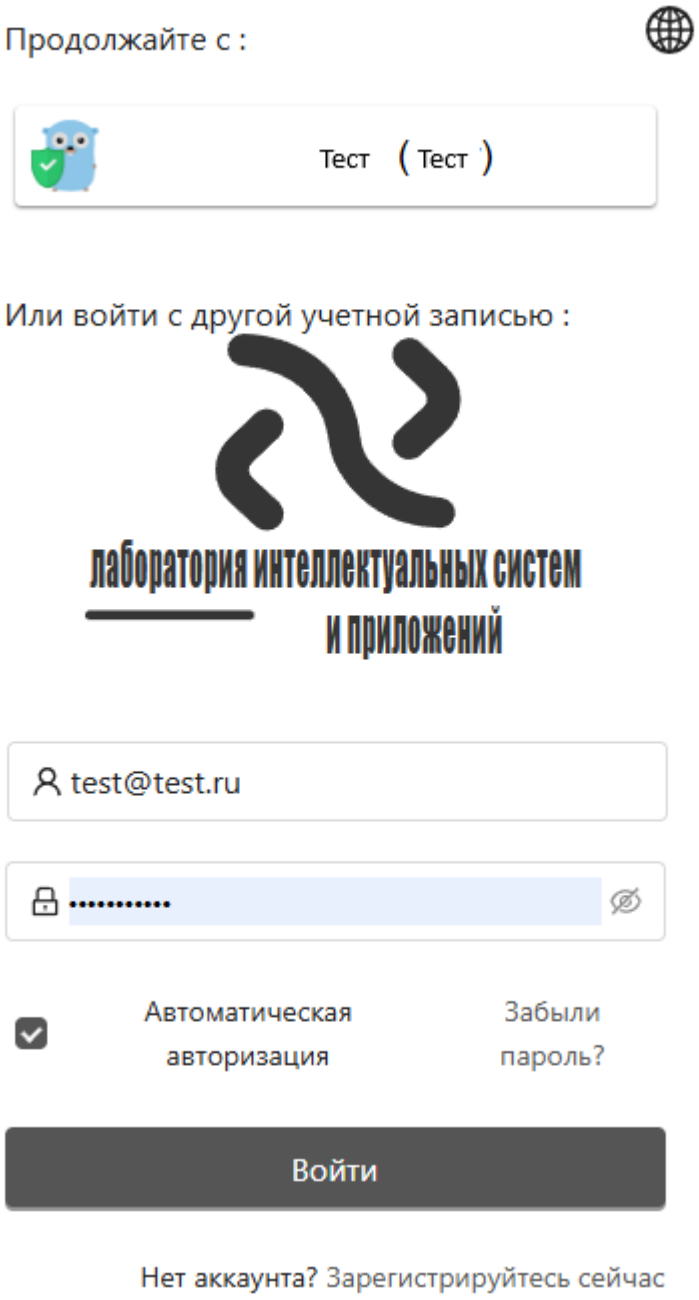


Рис. 6. Ввод учётных данных

2. Основное окно сервиса

На Рис. 7 представлен интерфейс программы для управления задачами. В левом верхнем углу находится меню с разделами:

- Задачи: основной раздел для управления задачами по сбору первичной информации.
- Ленты: раздел для работы с созданными лентами, состоящих из набора фактов.
- Факты: раздел для просмотра фактов, собранных с помощью задач.
- Поиск: инструмент для интеллектуального поиска информации в сети Интернет.
- Поддержка: в нижней части экрана находится кнопка «Поддержка», которая позволяет обратиться в службу поддержки.

В центральной части экрана находится рабочая область, вид которой зависит от выбранного пункта меню. В правом верхнем углу расположен переключатель языка сервиса и доступ в личный кабинет.

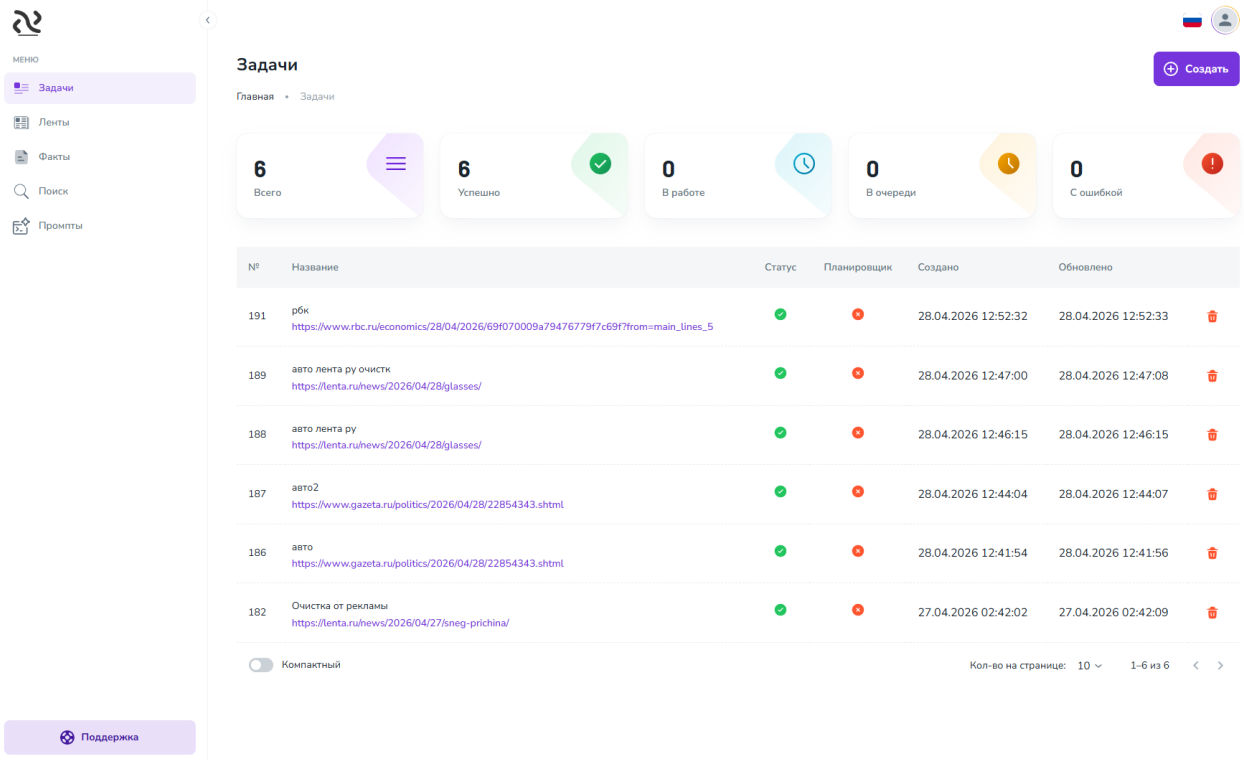


Рис. 7. Основное окно сервиса

3. Задачи

Вкладка «Задачи» отображает список всех задач, которые пользователь создал для сбора первичной информации. На основном экране расположена таблица, содержащая информацию о задачах, общая статистика по выполненным и выполняемым задачам и кнопка «Создать» для создания новой задачи. Пользователь может создать 5 типов задач:

- парсинг сайта с планировщиком или без,
- обработка файла, загруженного с локальной машины,
- парсинг каналов Telegram,
- выгрузка и последующая обработка файла с FTP сервера,
- получение дайджеста по заданной теме.

3.1. Создание задачи на сбор первичных данных с сайта

Данный тип задач предназначен для сбора информации с сайтов в сети Интернет (Рис. 8). Сборщик обрабатывает только текстовую информацию сайта.

Назначение полей:

- «Название» – название задачи которое будет отображаться в списке задач.
- «Сайт (URL)» - адрес страницы для парсинга.
- Переключатель «Без Обхода»/«В глубину» выбирает будет ли выполнятся парсинг одной страницы или будут обрабатываться также и ссылки, ведущие на другие страницы. При выборе «В глубину» дополнительно необходимо указать префикс сайта и глубину обработки ссылок. Префикс сайта указывает с какого сайта обрабатывается информация. Если ссылки ведут на другие сайты, то они не будут обработаны.
- «Планировщик» - устанавливает однократное или периодическое выполнение задачи (Рис. 9).
- «Использовать упрощенный сборщик» - Упрощённый сборщик обрабатывает только HTML страницы. Использование упрощённого сборщика повышает скорость обработки сайта.
- «Использовать очистку LLM» - полученная текстовая информация дополнительно обрабатывается для удаления из текста рекламных вставок, ссылок на другие сайты и аналогичного «шума».
- «Предсоздать ленту» - При включенном переключателе собранные факты автоматически группируются в ленту.
- «Выберите промпт для обработки ленты» - выбирается промпт для консультанта по материалам ленты, из списка ранее созданных (Рис. 23).

При использовании этой опции лента обрабатывается автоматически, после обработки лента помечается значком . Результат обработки доступен в специальном разделе ленты (Рис. 11).

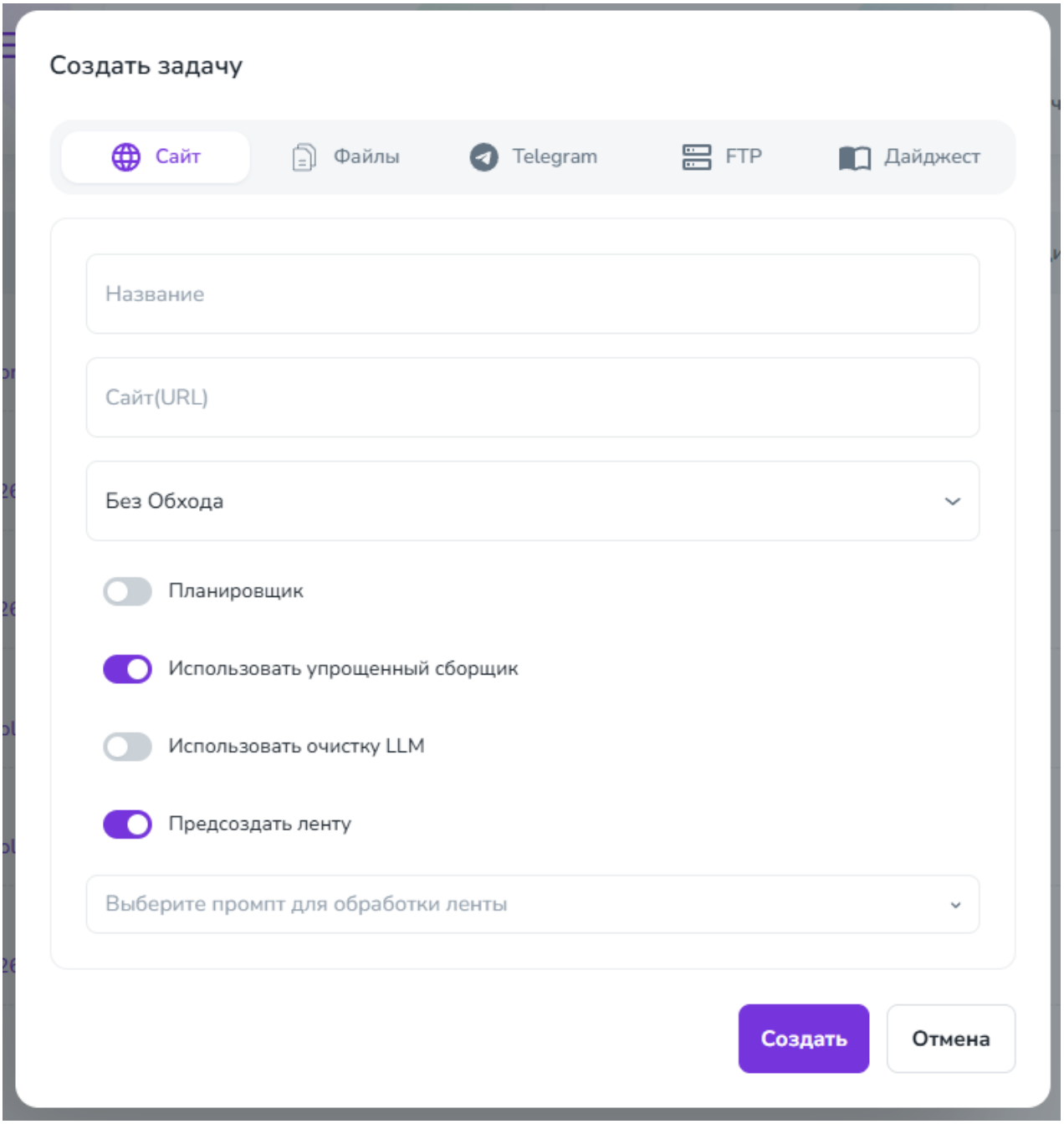


Рис. 8. Создание задачи на сбор информации с сайта

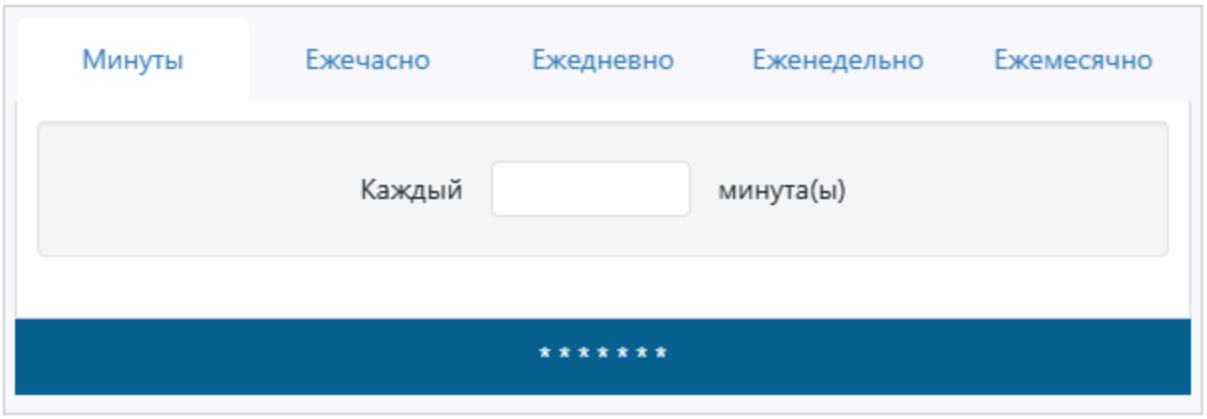


Рис. 9. Настройка планировщика

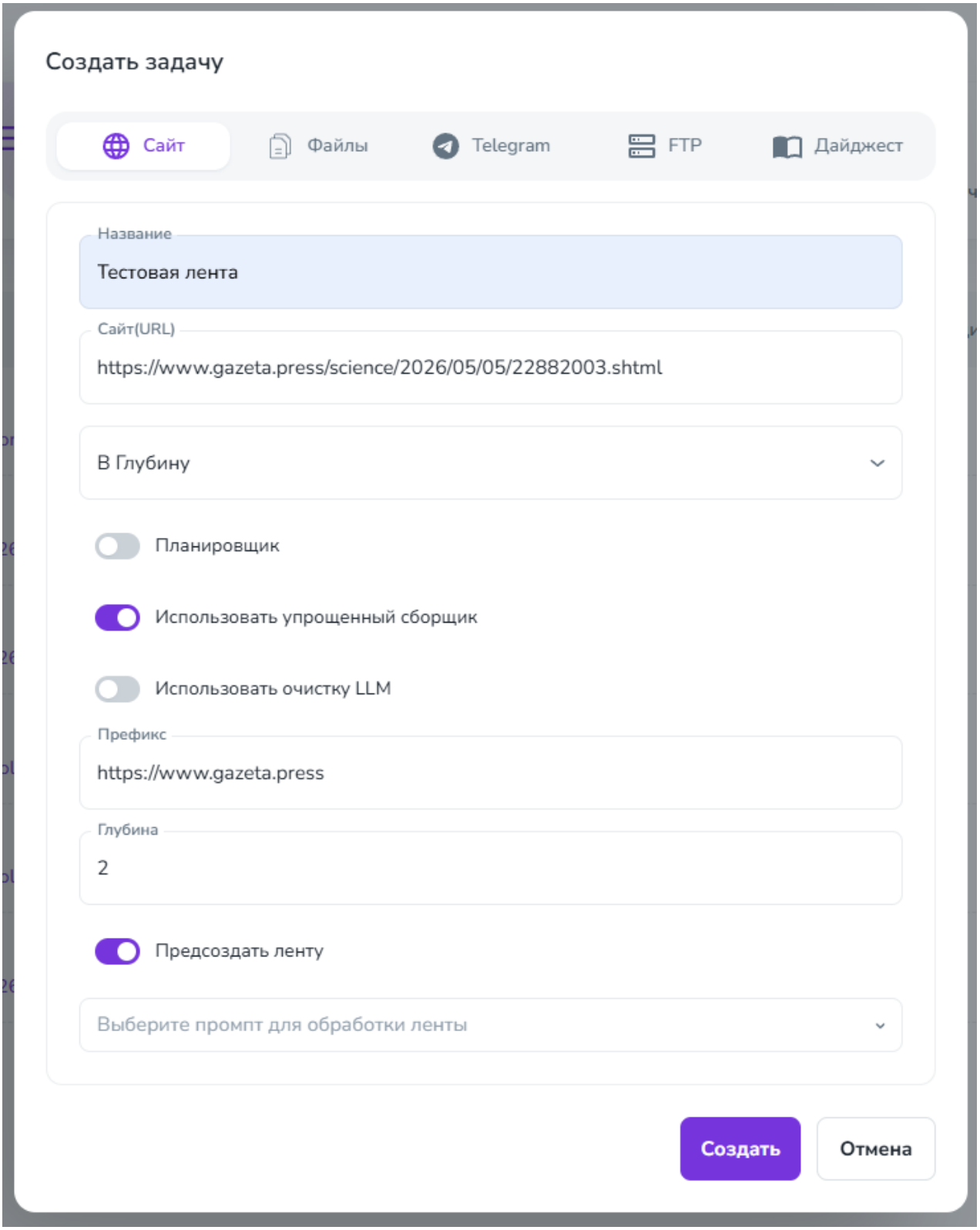


Рис. 10. Настройка сбора в глубину

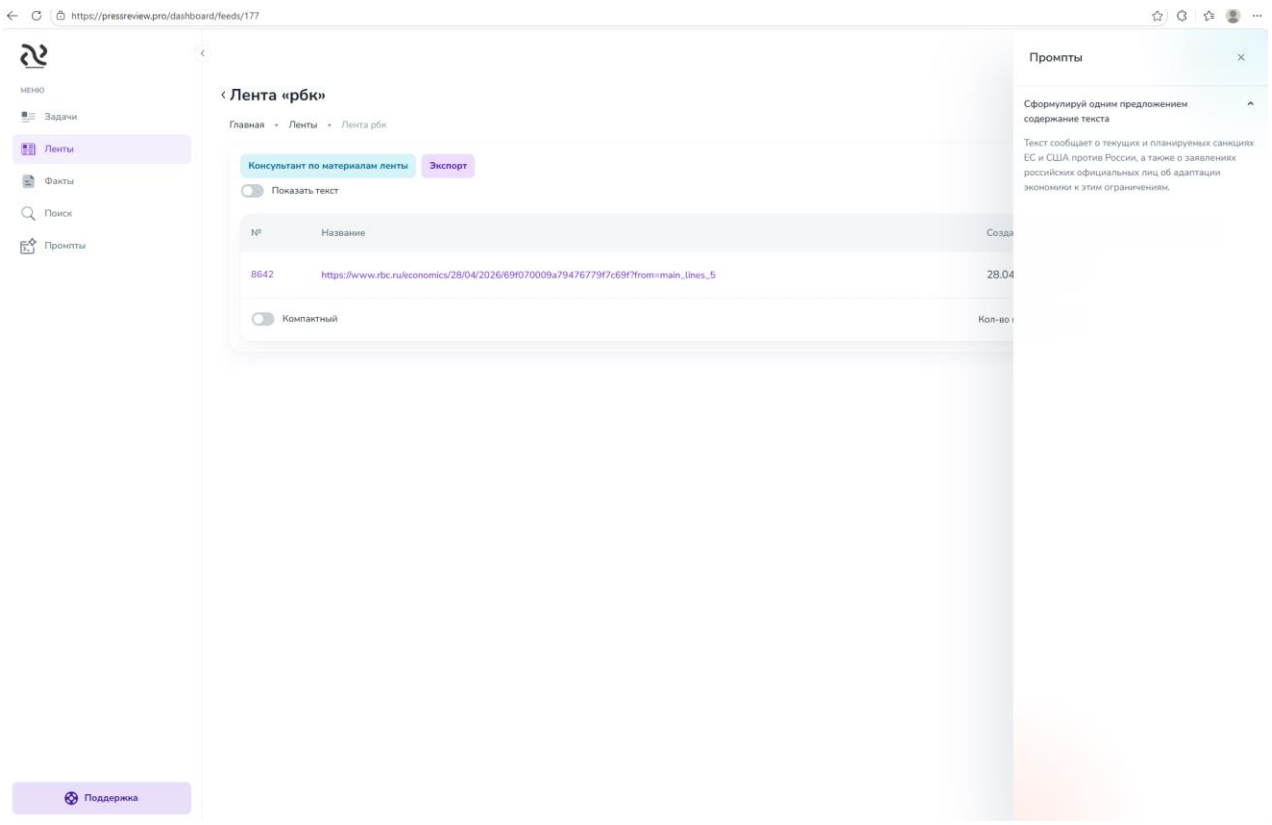


Рис. 11. Вывод результата автоматической обработки ленты

3.2. Создание задачи на сбор первичных данных из файла

Сервис позволяет извлекать текстовую информацию из файлов в форматах pdf, doc, docx, txt, xlsx, rtf, jpg, png. Загрузить файл можно через выбор в «проводнике» или при помощи технологии «перетаскивания» файла форму (технология drag and drop) (Рис. 12).

Назначение полей:

- «Название» – название задачи которое будет отображаться в списке задач.
- «Предсоздать ленту» - При включенном переключателе собранные факты автоматически группируются в ленту.

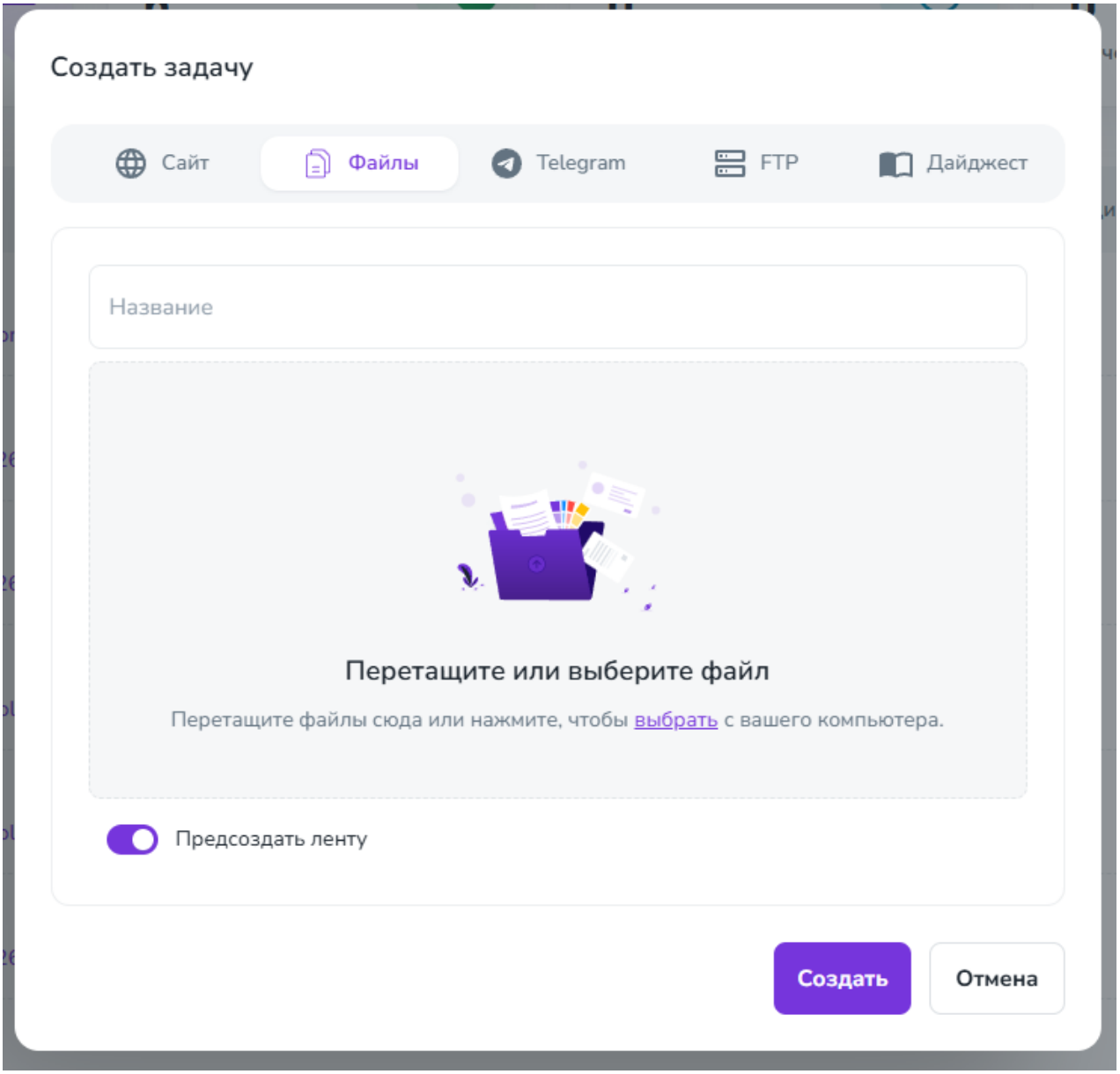


Рис. 12. Сбор текста из файла

3.3. Создание задачи на сбор первичных данных из канала Telegram

При помощи телеграм-бота, вы можете собирать информацию из телеграм-каналов (Рис. 13). Назначение полей:

- «Название» – название задачи которое будет отображаться в списке задач.

- Ссылка (URL) - ссылка на телеграм-канал (ссылка обязательно должна быть в формате https://t.me/link, иногда ссылки на телеграм начинаются с tg://link).

Количество - строка обозначает количество новостей, которые необходимо собрать. Не рекомендуется вводить слишком большое значение (больше 30), так как телеграм может блокировать за излишнюю активность.

- «Предсоздать ленту» - При включенном переключателе собранные факты автоматически группируются в ленту.

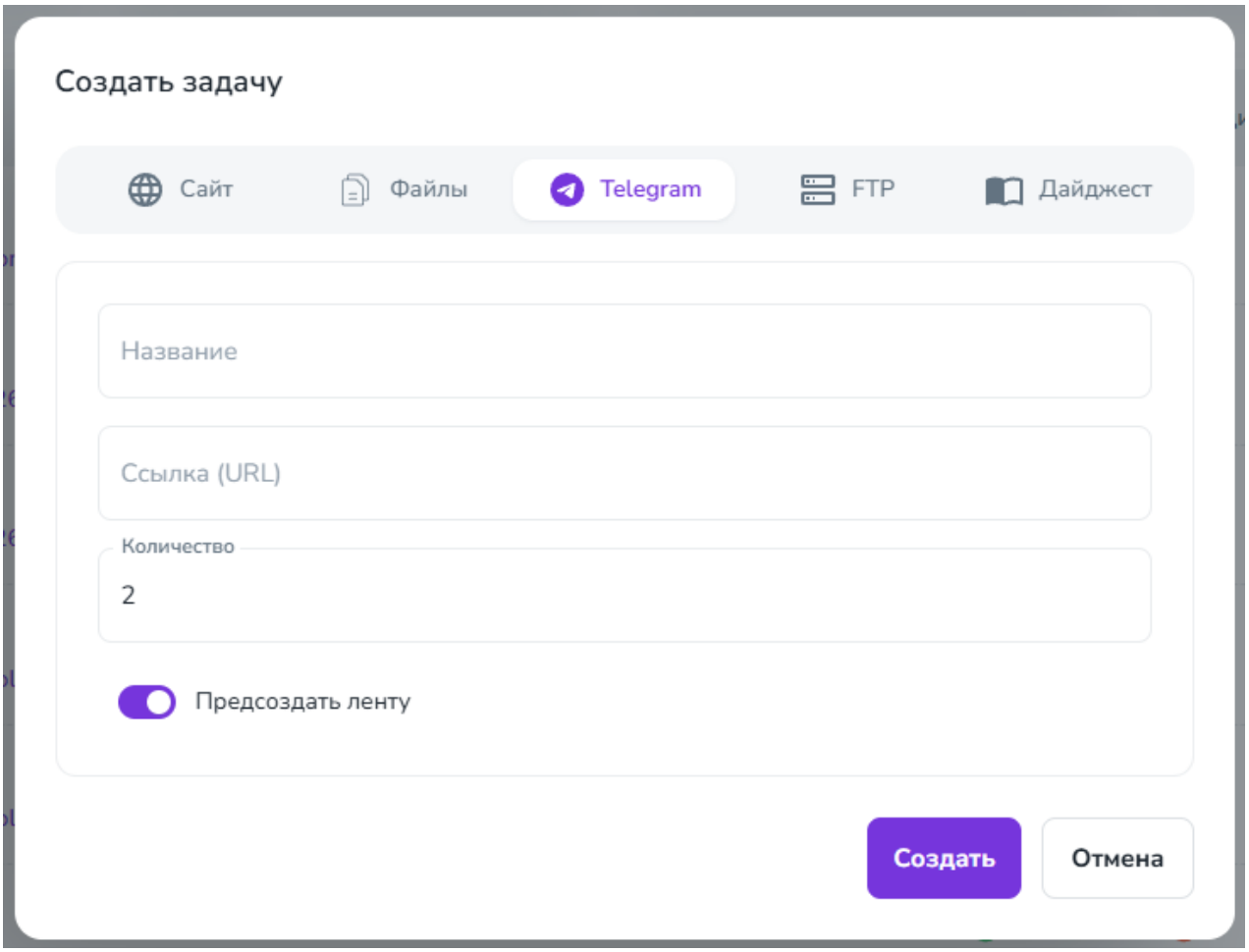


Рис. 13. Сбор данных из канала Telegram

3.4. Создание задачи на сбор первичных данных из файлов, находящихся на ftp сервере

Для выгрузки с FTP сервера вам необходимо иметь необходимые для доступа данные (адрес, логин, пароль).

После ввода данных, вы сможете просматривать список файлов на сервере и выбрать те, которые хотите выгрузить.

Важно! Если ftp сервер открыт и не имеет логина и пароля, в соответствующие поля вам нужно ввести логин: anonymous пароль: guest

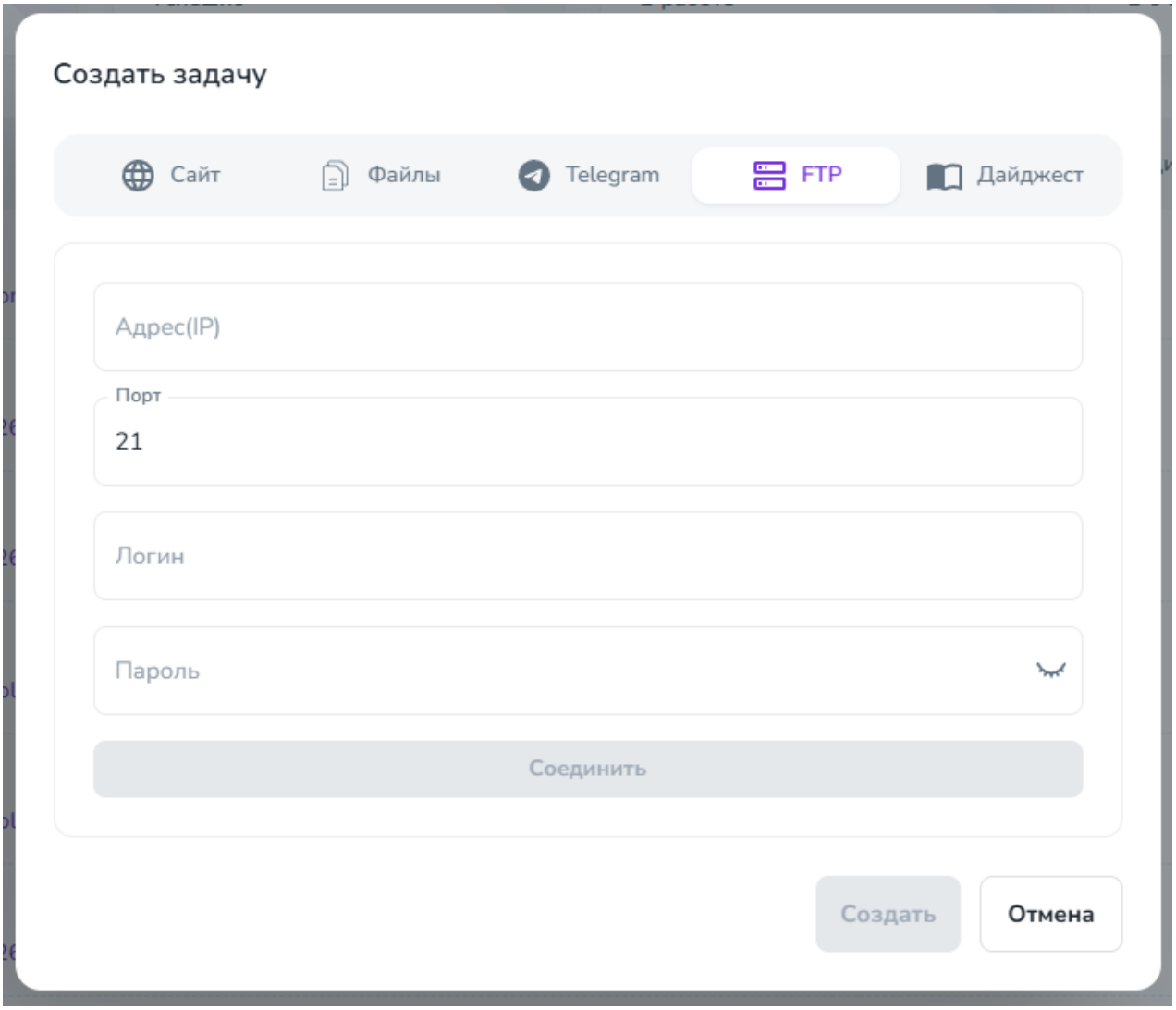


Рис. 14. Сбор данных из файлов, расположенных на ftp сервере

3.5. Создание задачи на получение дайджеста

Тип задачи «Дайджест» предназначен для получения краткой справки по заданной теме. Для формирования дайджеста необходимо указать тему, по которой формируется дайджест (Рис. *15*). В процессе выполнения задачи сервис выполняет в автоматическом режиме следующую последовательность действий:

- по заданной теме определяется набор ключевых слов, наиболее полно определяющих тему,
- на основании набора ключевых слов составляется поисковый запрос по теме,
- по сформированному запросу выполняется поиск информации на сайтах сети интернет,
- результаты выполнения запроса очищаются от рекламы и другого сопутствующего «мусора» и собираются в ленту,
- на основании полученных материалов составляется краткая справка по заданной теме,
- справка выводится в результатах обработки текста ИИ помощником ().

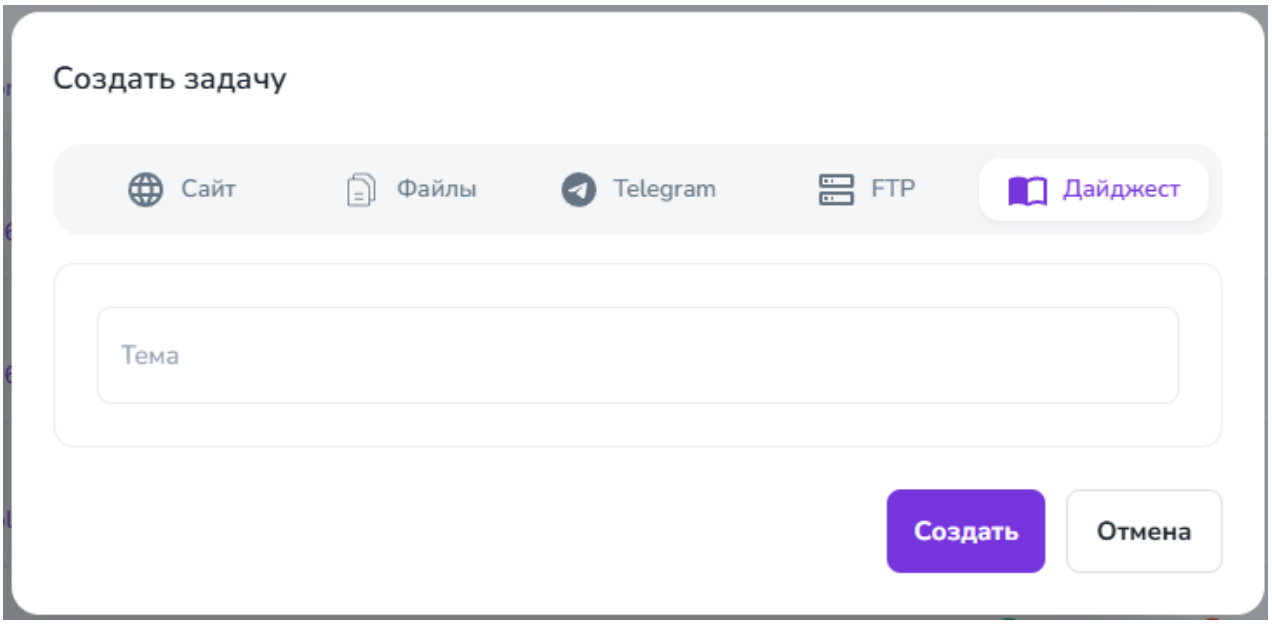


Рис. 15. Формирование дайджеста

4. Факты

Факты – это все документы, которые загружены при помощи сервиса Дайджест (Рис. 16). При обработке фактов сервис выделяет локации, персоны, организации и другие значимые понятия (Рис. 17).

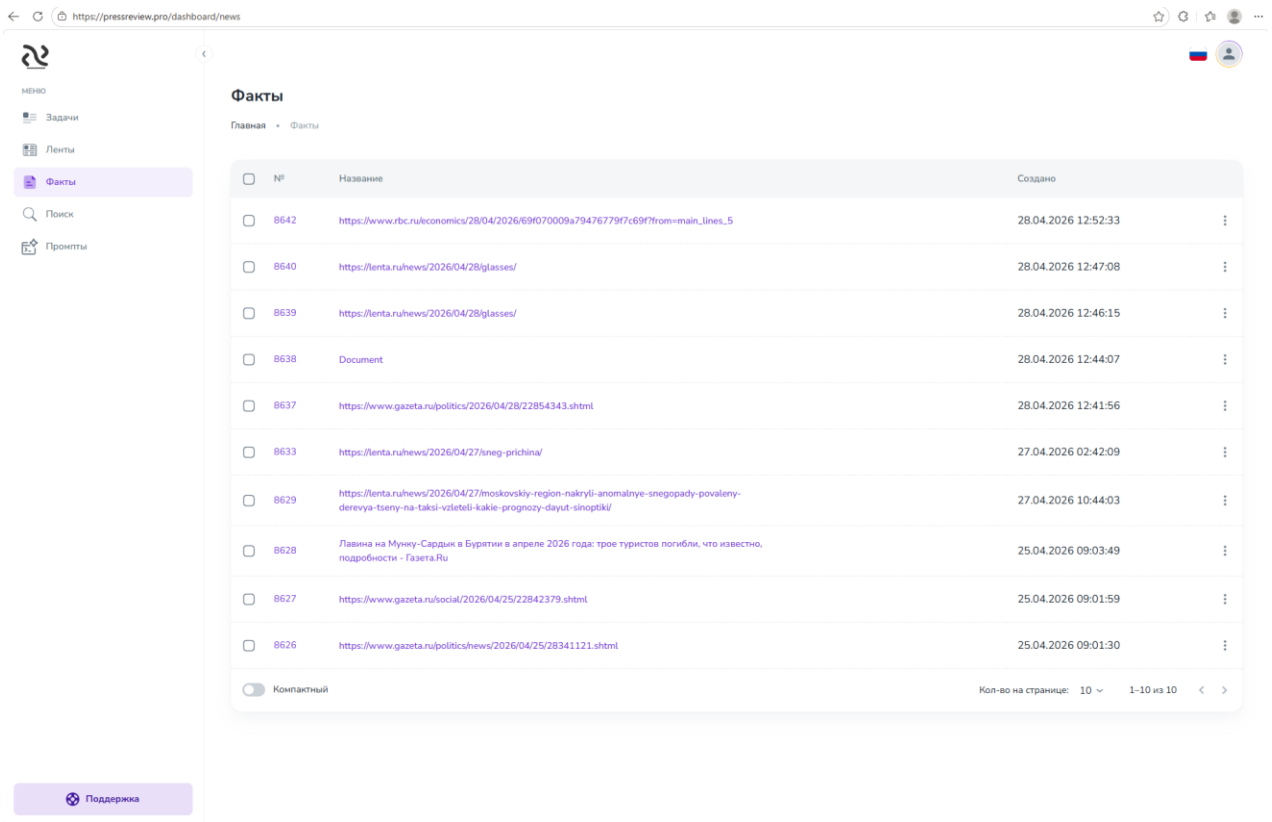


Рис. 16. Результат сбора фактов

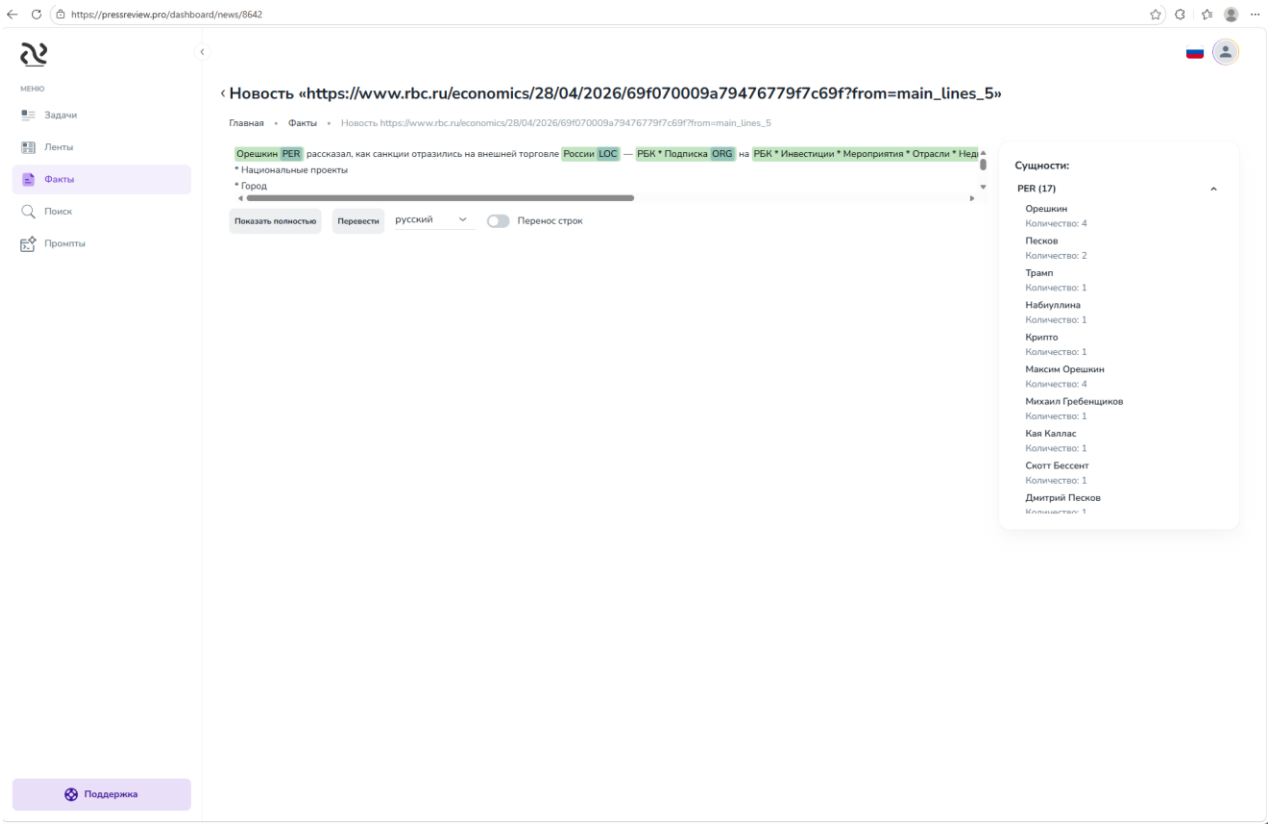


Рис. 17. Обработка фактов

5. Ленты

Раздел «Ленты» используется для группировки фактов по темам и последующей обработки собранной информации (Рис. 18).

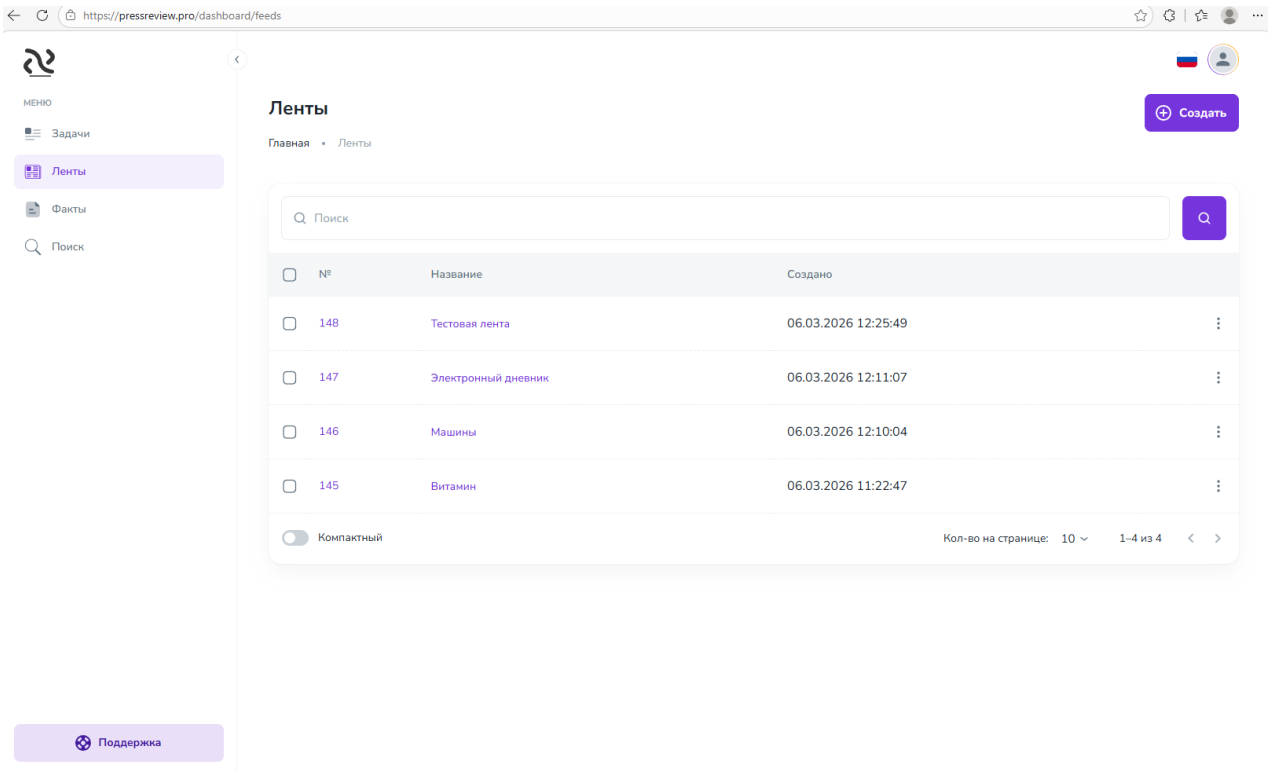


Рис. 18. Раздел «Ленты»

Лента может создаваться автоматически при постановке задачи на сбор информации (см. раздел Задачи) или вручную из уже собранных фактов. Для создания ленты вручную необходимо нажать кнопку «Создать». При нажатии на кнопку «Создать» откроется страница создания ленты (Рис. 19).

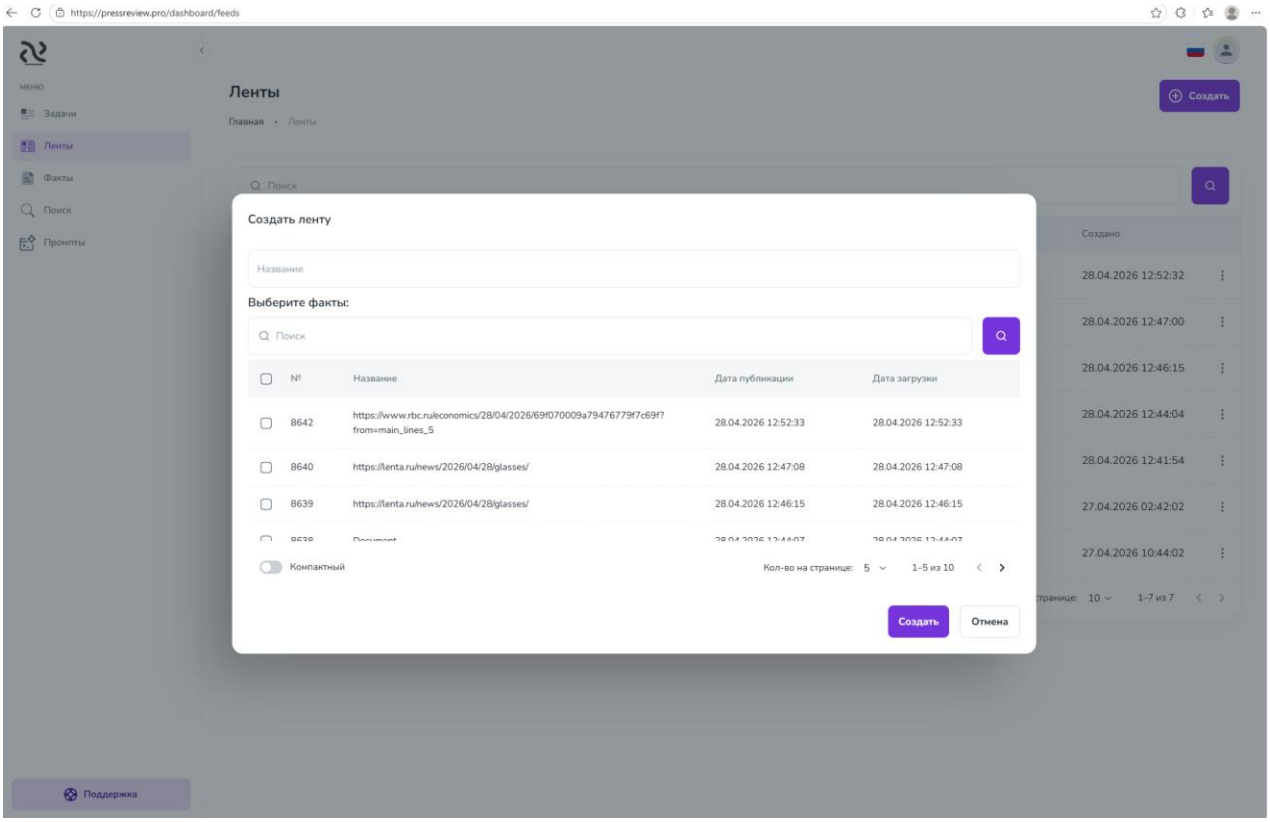


Рис. 19. *Создание Ленты*

Перечень полей и действий при создании Ленты:

- «Название» - название создаваемой ленты.
- «Поиск» - поиск по информации, содержащейся в фактах.
- Ниже расположена таблица с перечнем фактов, которые можно отметить для включения в Ленту.
- «Создать» - сохранение Ленты.
- «Отмена» - отмена создания Ленты.

Работа с лентой (Рис. 20):

- «Консультант по материалам ленты» - основной функционал работы с лентой (Рис. 21).

- «Экспорт» - экспорт материалов ленты в текстовый файл.

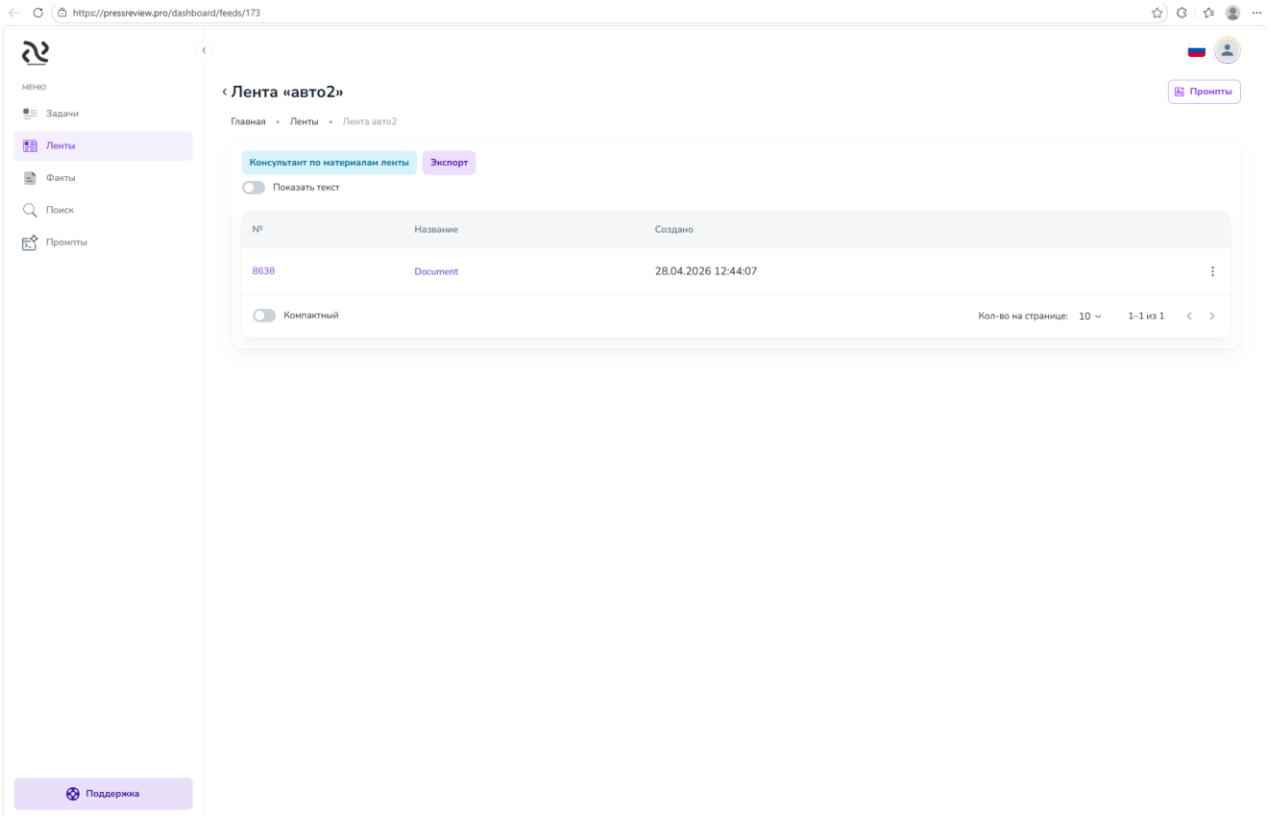


Рис. 20. Работа с лентой

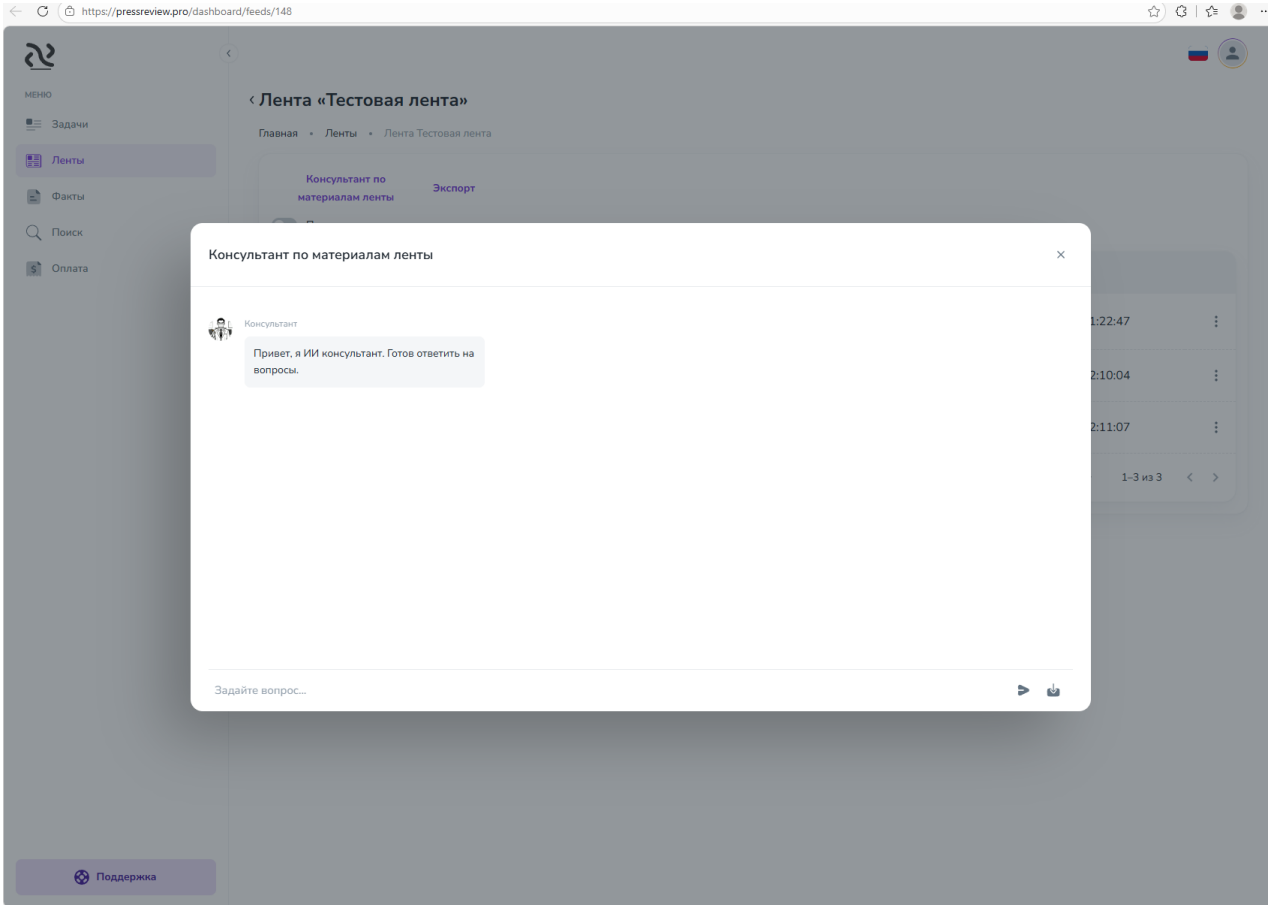


Рис. 21. ИИ консультант по работе с лентой

6. Поиск

Раздел «Поиск» позволяет собирать информацию из сети Интернет при помощи поисковых систем и собственных алгоритмов поиска информации (Рис. 22).

- Поле «Поисковый запрос» предназначено для ввода текста запроса.
- Переключатель «Учитывать близкие по смыслу выражения» включает функцию автоматического обогащения запроса словами и выражениями, которые семантически близки к запросу. Например, слово «Эпидемия» связано с следующими по смыслу словами: Болезнь, Вирус, Пандемия,

Инфекция, Карантин, Здравоохранение и другими. Запрос ведётся с учётом всех дополнительных понятий.

- Ползунок устанавливает «Строгость» соответствия результату поиска запросу. Чем больше значение ползунка, тем выше точность соответствия найденных материалов запросу.

Ниже отображается статус поиска:

- Успешный сбор - сайт успешно обработан
- Ошибка при сборе - не удалось обработать сайт
- Обработано - информация с сайта получена и обработана (после этого она выводится в разделе поиск)
- Скрыто - если результат поиска собран, но его температура не соответствует заданной пользователем, результат скрывается.

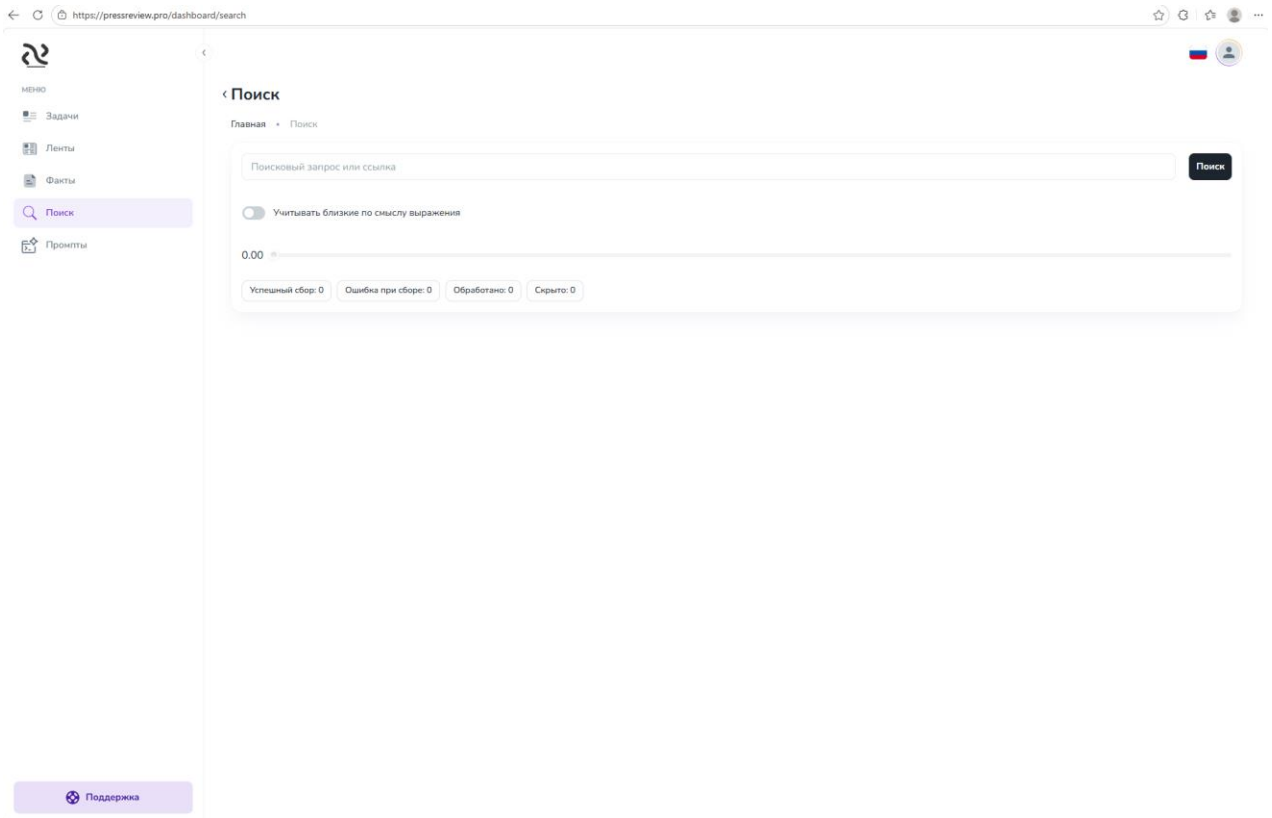


Рис. 22. Интеллектуальный поиск

7. Промпты

В этом разделе выполняется подготовка промптов для автоматической обработки загруженных текстов с помощью ИИ (Рис. *23*).

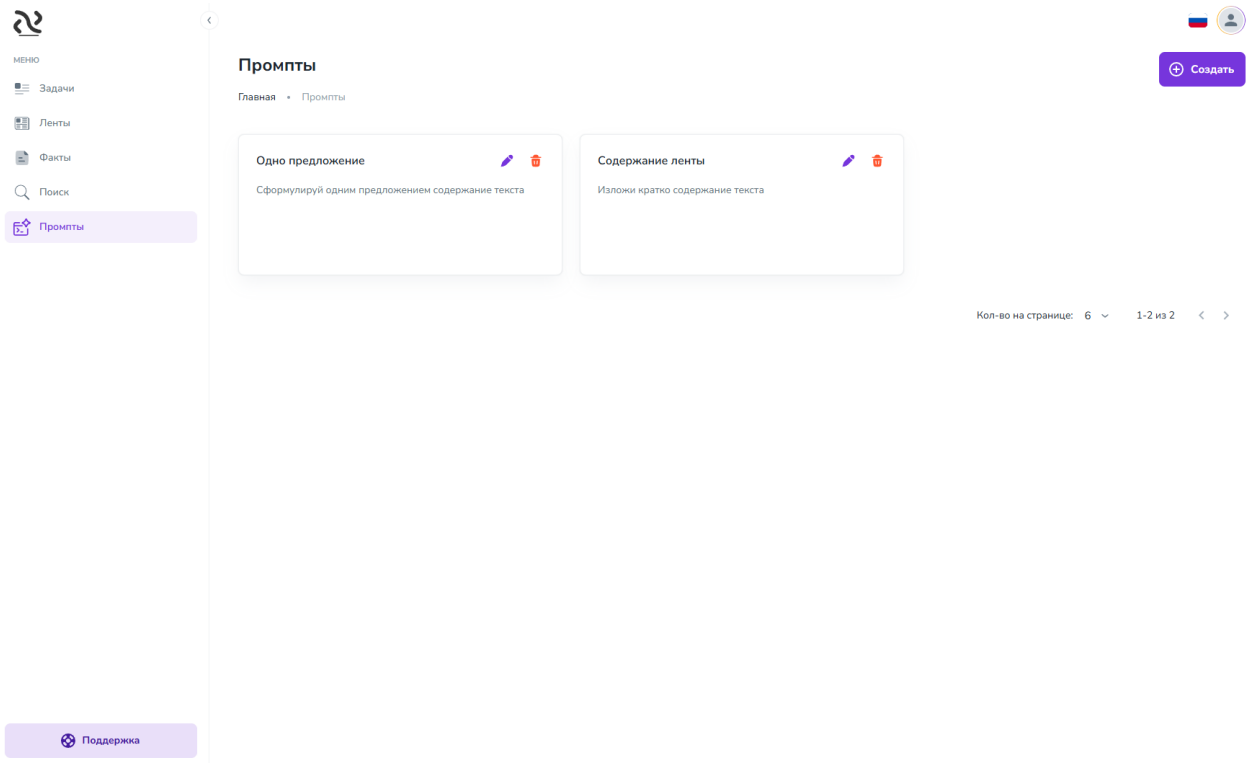


Рис. 23. Раздел «Промпты»

8. Поддержка

Раздел предназначен для обращения в службу поддержки (Рис. *24*). Ответ на обращение придёт на почту, указанную в обращении.

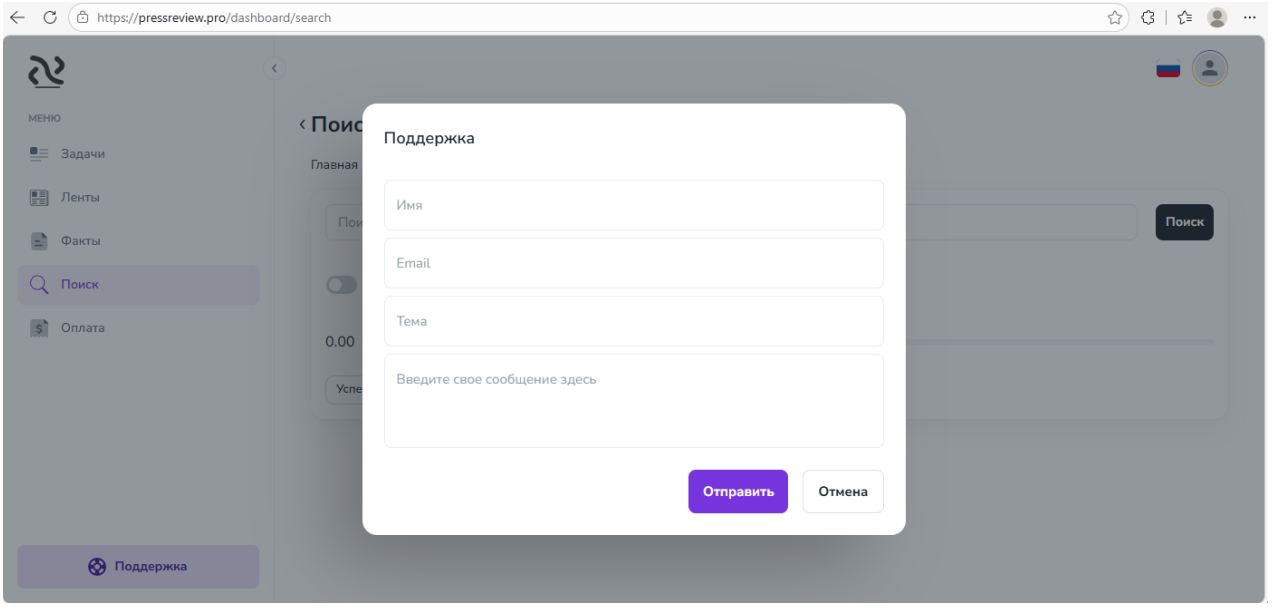


Рис. 24. Форма обращения в службу поддержки