ResetLab

ETL 4 AI

Отраслевые кейсы применения ИИ для обработки данных

Типовой RAG пайплайн

Сбор технических документов

Сбор документов для загрузки

Разбиение на чанки

Разделение документов на меньшие фрагменты

Индексация векторных представлений

Хранение векторных представлений в векторной базе данных

Создание векторного представления запроса

Преобразование запроса в векторное представление

Извлечение релевантных фрагментов

Извлечение релевантных фрагментов из базы данных

Генерация ответа LLM

Генерация ответа с помощью LLM

Предоставление ответа пользователю

Предоставление ответа пользователю

Предварительная обработка документов

Очистка и подготовка документов

Создание векторных представлений

Преобразование фрагментов в векторные представления

Пользовательский запрос

Пользователь вводит запрос

Поиск релевантных фрагментов

Поиск релевантных фрагментов в векторной базе данных

Формирование контекста

Создание контекста для LLM

Постобработка ответа

Очистка и форматирование ответа

ResetLab

Данные есть — пользы мало

Несъедобные большие данные

В архивах компаний хранится огромное количество данных но они не работают

Аудио и видео

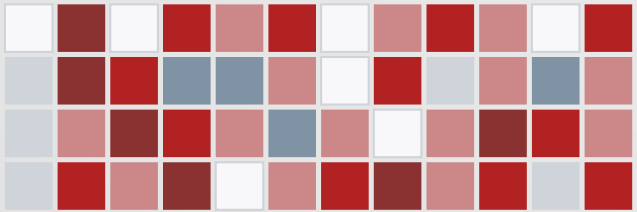
Неструктурированные аудио и видеофайлы

Текст

Макеты страниц, таблицы, формулы

Скан-копии

Сканированные документы, требующие обработки



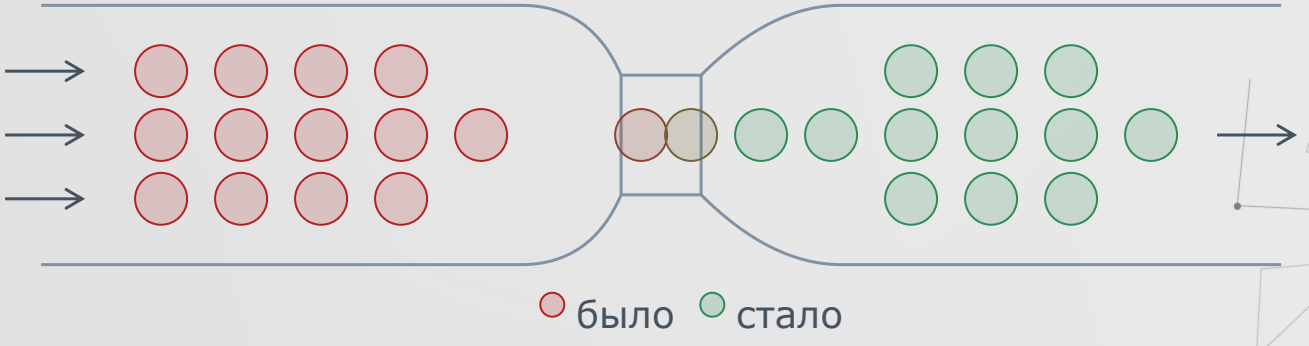
Нужно сделать AI-проекты быстрее и дешевле

Хотелось бы

- Сократить time-to-production
- Уменьшить стоимость подготовки данных
- Снизить зависимость от ручной работы

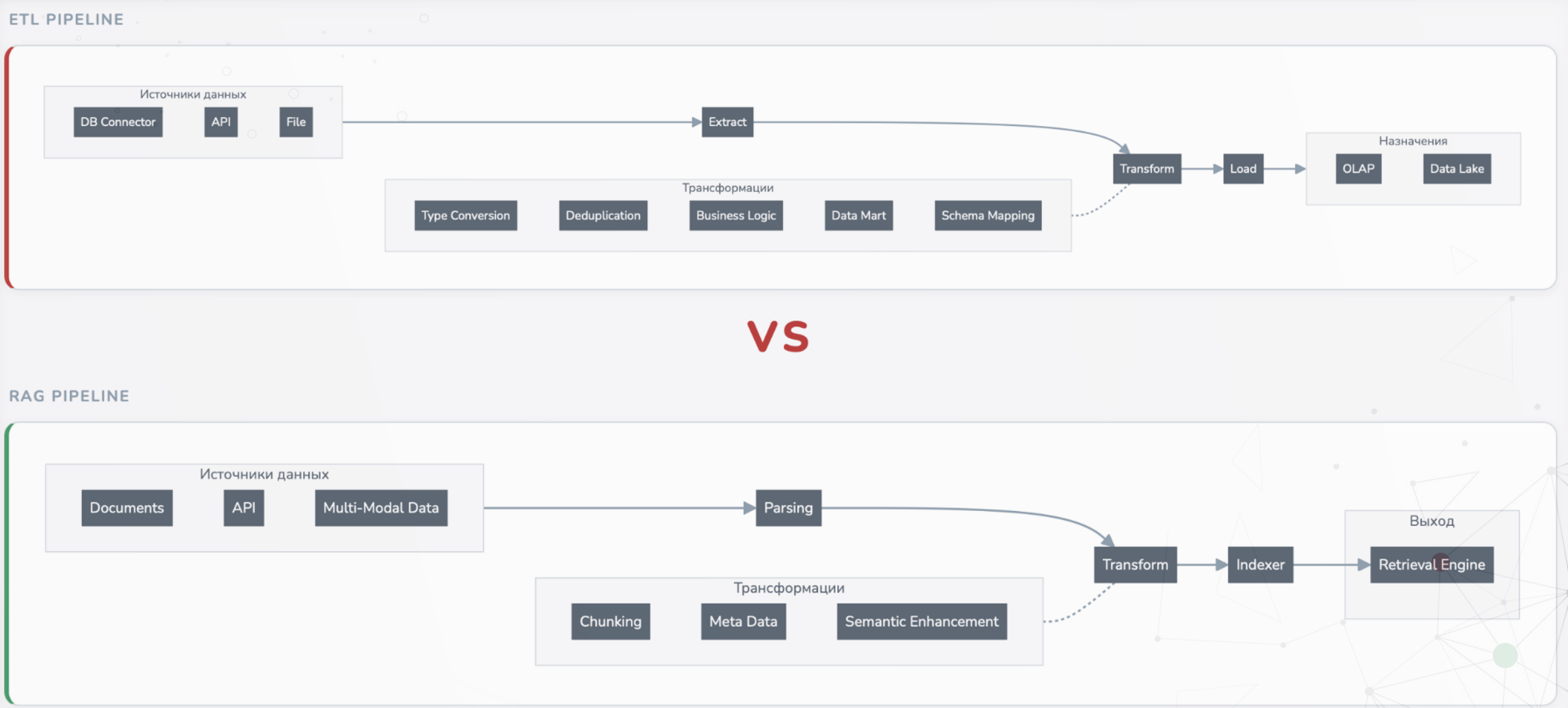
Ключевой фактор

- Качество результата сильно зависит от промптов
- Пользователи должны сами настраивать поведение системы



ResetLab

Сходство RAG–пайплайна с ETL/ELT



ResetLab

Многие проекты используют одни и те же кирпичи снова и снова

Парсинг
PDF · картинки · DOC · PPT · таблицы · прочее

Чанкинг
фиксированный · семантический · рекурсивный

LLM-трансформации
сводка · извлечение · классификация · прочее

Мэтчинг
эмбеддинги · косинусное сходство

Индексация (для RAG сценария)
гибридный поиск: BM25 + vector

Процессинг
Многопоточность · мониторинг · отладка

ВЫВОД

Нужен конфигурируемый пайплайн, а не набор скриптов

ResetLab

Что должна уметь система

Универсальный парсер:

- Word, Excel, PDF
- изображения (OCR, классификация)
- аудио / видео
- сложные документы (таблицы, формулы)

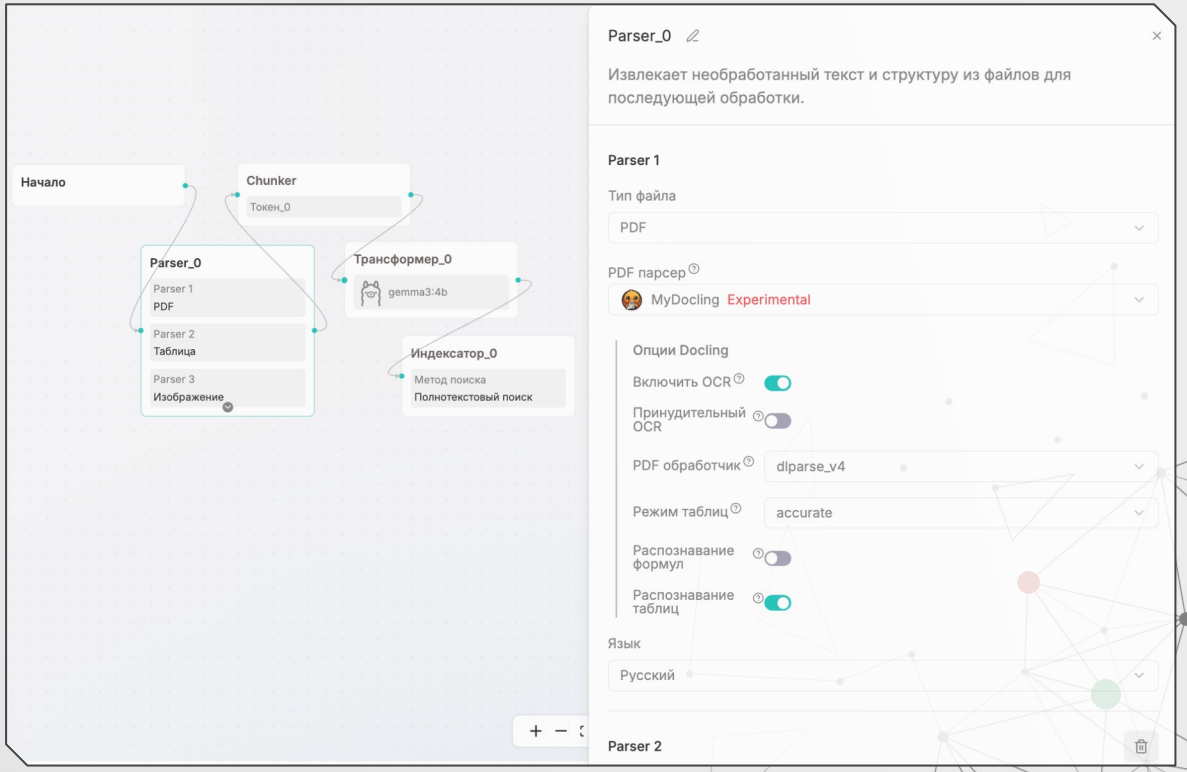
Масштабируемая обработка:

- batch-режим
- многопоточность
- быстрое масштабирование

Гибкая настройка:

- пайплайны
- промпты
- сценарии обработки

Пример простого пайплайна



Взяли основу — и сделали production-ready

Базовый проект

ragflow (Apache License 2.0, ⭐ 75k)

Что добавили

- 🇷🇺 Улучшили обработку русскоязычного текста
- 🧩 Добавили парсеры и интеграции (Yandex GPT, Yandex Search, etc)
- 🧱 Разработали дополнительные узлы обработки (BI-агент, мэтчинг)
- 🛒 Проработали библиотеку кейсов использования
- 📋 Подготовили дорожную карту дальнейших улучшений

Результаты

RAGu

Получилась платформа для **быстрой сборки AI-решений**

Бизнес доволен

ResetLab

Интеграция

Компании и пользователи

Экосистема провайдеров моделей

Grok, Claude, Mistral, Perplexity, deepseek, Gemini

Хранилище данных

Отчеты и презентации

Приложения и системы

Прогнозные модели

Оптимизационные модели

DB коннект

Индексация

API вызов

Агенты

Exe SQL

AI Поиск

База знаний

Рекомендации

Прогнозы

Сигналы

Аналитика

Отчеты

Встроенный чат

MAX

Telegram

API

AI Визуализация

AI ассистенты

RAGu использует API приложений и СУБД как инструменты взаимодействия с ресурсами заказчика для генерации достоверных и обоснованных ответов AI ассистента

ResetLab

Анализ объявлений для надзорных служб мегаполиса

Потенциалы экономии

1. Трудоемкий анализ публичных источников информации
2. Многообразие нормативной документации видов разрешенной деятельности (ВРИ)

Решение

1. Консолидация нормативной документации по ВРИ: построена специальная база знаний (RAG)
2. Реализован гибридный механизм поиска канонических формулировок ВРИ
3. Встроена компактная нейронная сеть для анализа публичных объявлений в режиме потоковой обработки

Результаты

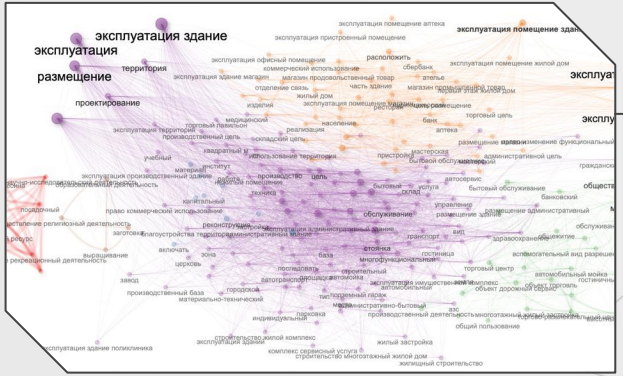
1. Интеграция с ГИС системой заказчика
2. Генерация отчетности для трассировки результатов разбора контента
3. Все компоненты размещены на слабом сервере заказчика без GPU ускорителя
4. Сокращение времени на поиск нарушений 90%

python

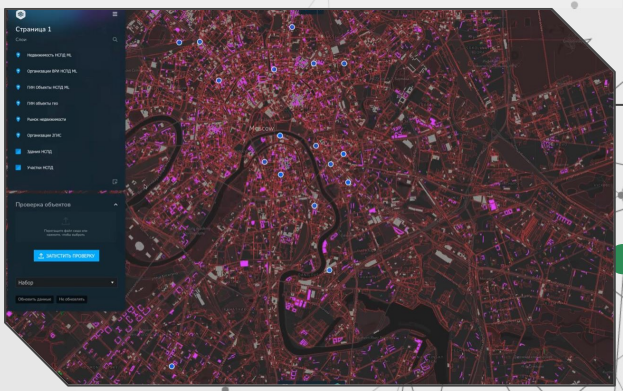
docker

RAGu

Семантический анализ текста



Интеграция в приложение заказчика



ResetLab

Обработка технических паспортов для энергетики

Потенциалы экономии

1. Ручное извлечение тех. Параметров из 12 тыс. документов
2. Целевые сведения хранятся в таблицах, формулах, рукописном тексте

Решение

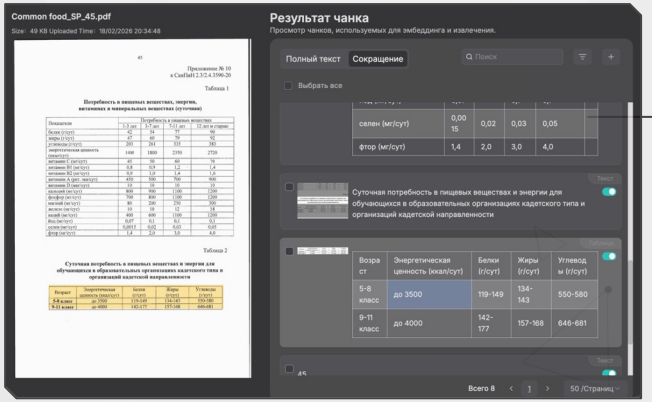
1. Подбор комбинации OCR для получения текстовых сведений там где даже человеку сложно разобрать
2. Извлечение тех. параметров LLM
3. Мэтчинг со справочником тех. Параметров с помощью гибридного поиска

Результаты

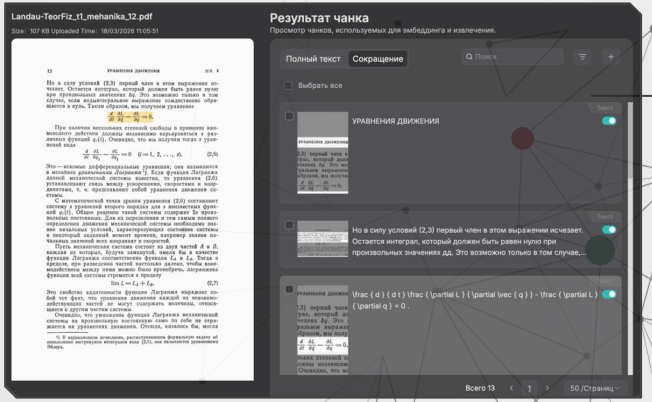
1. Сокращение ручного труда на 95%
2. Единообразный результат, готовый к загрузке в систему управления производственными фондами

RAGu

Обработка таблиц



Обработка формул



ResetLab

СПАСИБО

Вопросы?

www. resetlab.ru
d.kibalnikov@resetlab.ru
+79119714950

@KB_DM

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.