

# Тестовое задание: "Каталог данных и парсер сложных форматов"

Цель: Разработать сервис для сбора, анализа и хранения метаданных с поддержкой асинхронных задач, API и сложными форматами данных.

## Основные задачи:

### Импорт и парсинг сложных данных:

- Разработать модуль для парсинга метаданных из многослойных файлов (JSON, XML).
- Поддерживать динамическое извлечение схем данных с автоматическим определением типов колонок.

### Хранилище метаданных:

- Построить схему базы данных для хранения метаданных, включая версии схем, связи между наборами данных и историю изменений.
- Реализовать механизм автоматического создания схем данных из парсинга.

### Асинхронная обработка данных:

- Использовать Celery/Kafka для обработки больших файлов и периодических задач.

### Поиск и фильтрация:

- Реализовать контекстный поиск с использованием синонимов и нечётких совпадений.
- Добавить поддержку сложных предикатов и фильтров.

### API для работы с метаданными:

- Реализовать REST API с нестандартными методами ("MUTATE", "INTERROGATE", "REFINE").

### UI для отображение информации (на базе Bootstrap)

## Дополнительно:

### Визуализация связей:

- Построить граф связей между наборами данных с использованием react-flow или d3.js.

---

## Критерии оценки

- Качество кода (чистота, структура, PEP8).
- Оптимальность решений (выбор алгоритмов, запросов к БД).
- Масштабируемость (архитектура, использование очередей).
- Документация (описание API, примеры запросов).

---

## Как сдавать?

Репозиторий на GitHub/GitLab с исходным кодом.

Документация

Примеры запросов (curl/Postman)

Краткий отчет о принятых решениях.

Срок выполнения: 5–7 дней.

---

## Доп. файлы:

complex\_data\_with\_noise.xml

complex\_data\_final.json

final\_complex\_data\_with\_noise\_with\_errors.xml