
DocFlow Content Analyzer — поиск и классификация документов по содержанию

Datacult:

Денис Власюк - сбор и анализ информации

Яна Володина — занималась сайтом и frontend разработкой

Вячеслав Камлин — backend разработчик, логика кода



Цель:

Создать приложение
которое оптимизирует
работу с
многотысячными базами
данных

Задачи:

- 1) преобразовать pdf файлы в текст
- 2) построить иерархическую модель хранения файлов
- 3) классифицировать при помощи nlp файлы на категории и подкатегории
- 4) апробировать данное приложение

План работы:

- определение цели и задач;
- работа с источниками;
- создание логики кода;
- проработка отцифровки pdf файлов;
- кластеризация данных;
- написание frontend части проекта;
- создание иерархической системы для отображения файлов;
- создание поисковика.

Описание сценария работы с приложением

