

# Бриф

## Цели

1. Создать интерактивную базу знаний с продвинутым поиском (по датам, темам, авторам).
2. Создать чат-бота, консультирующего на основе данных, содержащихся в базе знаний. Данные в базе должны быть отранжированы в зависимости от их достоверности:
  - a. научная информация
  - b. научно-популярная информация
  - c. информация практиков
  - d. информация инфобизнесменов, эзотериков и т. д.
3. Создать инструмент для непрерывного блоггинга: (полу-)автоматическое формирование тематических материалов, подборок (литературы, курсов, фильмов и т. п.), обзоры (книг, курсов, фильмов и т. п.).

## Исходные данные

### Место хранения

Облако Яндекс.Диск и/или сетевой диск в локальной сети.

### Объём

Несколько ТБ. Материалы постоянно добавляются. Количество файлов — десятки (возможно, сотни) тысяч.

### Типы материалов

#### Видео

Нехудожественные фильмы (иногда многосерийные), видеогайды, учебные курсы (иногда из множества уроков, со смешанными материалами — не только видео), образовательные мероприятия (лекции, круглые столы, семинары, конференции и т. п.).

Форматы различные. Языки разные. Преимущественно русский.

#### Аудио

Аудиокниги, аудиогайды, настройки, учебные курсы (иногда из множества уроков, со смешанными материалами — не только аудио), образовательные материалы (запись лекций, круглых столов, семинаров, конференций и т. п.).

Форматы различные. Языки разные. Преимущественно русский.

#### Графика

Сканы печатных изданий.

Форматы различные. Языки разные. Преимущественно русский.

#### Текст

Печатные издания (книги, журналы, гайды и пр.), транскрипции, статьи, посты.

Форматы различные. Языки разные. Преимущественно русский.

## Последовательность обработки

Исходники				
Действие	Видео	Аудио	Графика	Текст
Результат				
Выделение аудиодорожки	<b>Аудиофайл</b> из видеофайла			
Транскрибирование	<b>Текст</b> из аудиофайла (желательно с таймкодами)	<b>Текст</b> из аудиофайла (желательно с таймкодами)		
Распознавание текста OCR			<b>Текст</b> из сканов	
Перевод текстов на русский	<b>Перевод</b> на русский	<b>Перевод</b> на русский	<b>Перевод</b> на русский	<b>Перевод</b> на русский
Озвучивание/титры	Видео с русской <b>озвучкой</b> и/или титрами	Аудио на русском языке		
Подготовка саммари	<b>Краткое содержание</b> основного текста	<b>Краткое содержание</b> основного текста	<b>Краткое содержание</b> основного текста	<b>Краткое содержание</b> основного текста
Обучение ИИ	<b>Чат-бот</b> , обученный на основных текстах			