



Первый этап создания ИС Мониторинг информационных ресурсов

ГЛАВНЫЙ РАДИОЧАСТОТНЫЙ ЦЕНТР

Рост аудитории социальных сетей, форумов, сервисов мгновенного обмена сообщениями

Значительное увеличение объемов информации в СМИ и сети Интернет

Сложность выявления большого объема ЗИ в «ручном» режиме.

Целенаправленное массированное распространение нарушений (медиакампании)

Постоянное появление новых источников и каналов распространения ЗИ

Тиражирование ЗИ на сайтах-зеркала

Распространение неточных (смысловых) копий ЗИ






Применение средств автоматизации без использования технологий ИИ не позволяет решать весь спектр задач выявления ЗИ.

Классические решения в области ИИ требуют значительных объемов обучающих выборок и временных затрат на обучение моделей

Необходимо использовать ИИ-решения на основе семантического анализа текста для выявления признаков нарушений в СМИ и сети Интернет














ИС МИР-1



	Цель:	Автоматизированный мониторинг информационных ресурсов сети «Интернет» и выявление информации, распространение которой в Российской Федерации ограничено или запрещено
	Исполнитель:	ООО «Вектор Икс» (ТВЦ) + МФТИ
	Руководитель проекта:	А.В. Федотов / НТЦ
	Бюджет / цена договора:	58,5 млн руб. / 57,8 млн руб.
	Даты подписания договора/акта:	30 сентября 2021 / 25 января 2022

Состояние ИС ГРЧЦ на момент старта проекта



	АС МАВР	АС МСМК	АС МАСМ	АС ЧИ
 ИИ	 Текст (глубокое ML)		<div style="border: 1px solid black; background-color: #f9e79f; padding: 10px; text-align: center;"> ЕДИНЫЙ МОДУЛЬ АНАЛИЗА  Текст (глубокое ML) </div>	
Поиск зеркал	-	-	-	-
 Анализ на словарях	 Только текст	 Только текст	 Только текст	-
 Сбор данных	 API IMDB, Кинопоиск	 Собственный WEB и API краулер	 Собственный WEB парсер	-
Поиск ресурсов с информацией	Вручную	Вручную	Вручную	 API Яндекс

Первый этап создания ИС МИР



		АС МАВР	АС МАСМ	АС МСМК	АС ЧИ	
 ИИ Глубокое ML Семантический анализ		 Только текст	ЕДИНЫЙ МОДУЛЬ АНАЛИЗА			 Только текст
			 Только текст	 Только текст	 Только текст	 Только текст
Поиск зеркал		-	Точные и неточные копии			Точные и неточные копии
 Анализ на словарях		 Только текст	 Только текст	 Только текст	-	
 Сбор данных		 API IMDB, Кинопоиск	 Собственный WEB и API краулер	 Собственный WEB парсер	 WEB и API краулер	
Поиск ресурсов с информацией		Вручную	Вручную	Вручную	 API Яндекс	

Первый этап создания ИС МИР



Функции ИС МИР-1

- Краулер
- Сервисы семантического анализа
- Аналитические сервисы
- Графический пользовательский интерфейс
- Аналитическая отчётность
- Средства технического мониторинга и администрирования

Технические характеристики

- Количество зарегистрированных пользователей не менее 10 000
- Количество одновременно работающих пользователей не менее 500
- Сервисы реализованы на основе независимых Docker контейнеров
- Применяется конвейер разработки (pipeline) для всех сервисов
- Управление данными на основе ETL (извлечение, преобразование, загрузка данных)

Функции сервисов семантического анализа

- Быстрое добавление новых видов нарушений (необходима небольшая обучающая выборка);
- Выявление нарушений близких по «смыслу» заданным
- Выявление именованных сущностей и отношений между ними
- Определение тональности (сентимента)
- Классификация медиматериалов по темам
- Хронологический анализ развития заданной темы
- Выявление неточных (смысловых) копий

Технические характеристики

- Классификация сентимента по F1 не менее 0.9
- Извлечение именованных сущностей по F1 не менее 0.92
- Идентификация именованных сущностей: полнота не ниже 0.75, точность не ниже 0.5
- Точность тематического моделирования не ниже 70%
- 10 000+ текстовых документов в минуту*

Требуется от 2 серверов с 4 производительными видеокартами в каждом

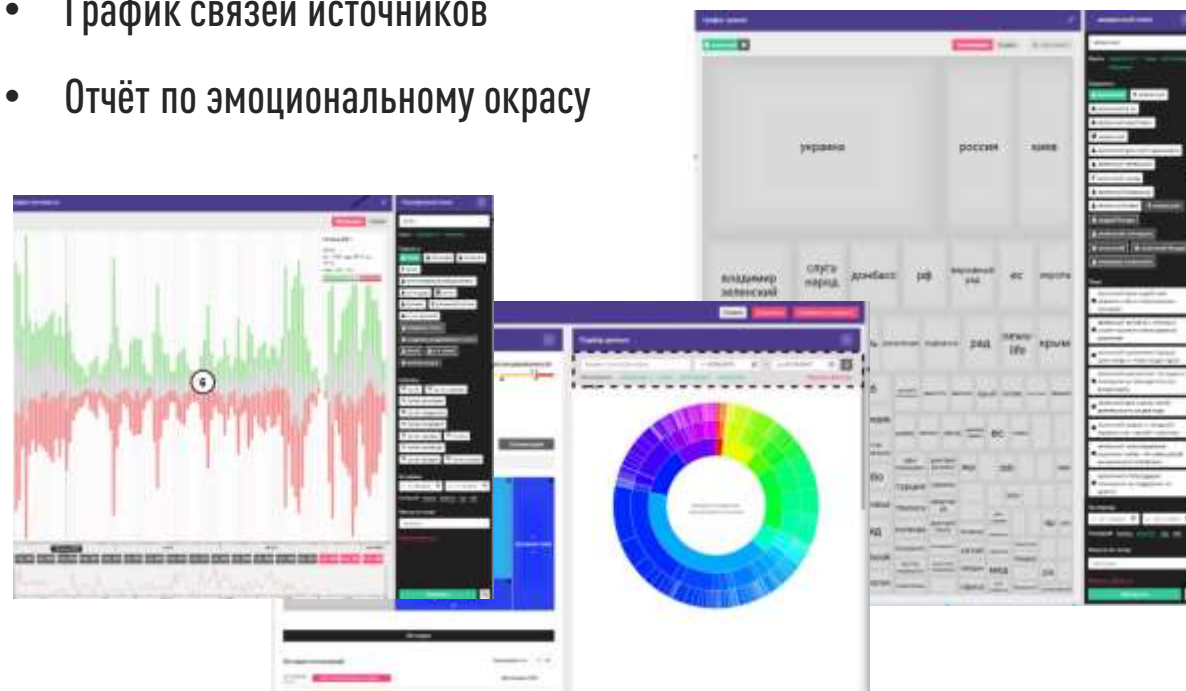
Аналитические сервисы

Выявление точных и неточных семантических копий

- Применение сервиса семантического анализа для поиска точных и неточных копий медиаматериалов
- Сервис поиска документов источников и обнаружения цепочки наследования медиаматериалов
- Выявление наиболее вероятных путей распространения копий медиаматериалов

Аналитическая отчетность

- График трендов
- График распространения новостей
- Тематическая картина медиапространства (тримап)
- График связей источников
- Отчёт по эмоциональному окрасу



Планы создания ИС МИР на 2022 год



Первый этап создания ИС МИР:

1. Закупка необходимого оборудования – март 2022
2. Развертывание ИС МИР – апрель 2022
3. Настройка и обучение ИС МИР на текущих датасетах – апрель 2022
4. Интеграция сервисов ИС МИР с системами направления СМК – июнь 2022
5. Опытная эксплуатация ИС МИР – сентябрь 2022
6. Промышленная эксплуатация – с октября 2022

2023 - 2024 годы

1. Консолидация информационных систем СМК
2. Создание подсистемы единого личного кабинета операторов (графический интерфейс и бизнес-логика)
3. Создание подсистемы информационной безопасности
4. Развитие краулера для анализа новых видов соцсетей и интернет-сервисов

5. Развитие сервисов ЕМА:
 - Сервисы анализа изображений и видео (Окулус)
 - Сервис автоматизации машинного обучения (Auto-ML)
 - Сервис анализа структурированных документов (doc, xls, pdf, odf и др.)
6. Повышение качества и точности сервисов

Потребность в иных подсистемах и сервисах будет выявлена по результатам работ первого этапа.

Целевая архитектура ИС МИР



Методы ИИ

Семантический анализ и онтология

Компьютерное зрение

Распознавание аудио

Поведенческий анализ

Выявление зеркал и фишинговых сайтов

Анализ источников

Анализ аудитории

Проверка фактов

Риск-ориентированный подход

Прогнозирование

Выявление манипуляций

Объяснимый ИИ (XAI)

Управление моделями ИИ

Управление разработкой

Управление версиями

Управление точностью

Управление качеством

Управление данными

Автоматизированная разметка

Ручная разметка

Безопасность данных

Репозитории данных

Хранилище данных (ETL)

Спасибо за внимание!