



ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК
ПРОГРАММНОГО КОМПЛЕКСА «МИП»

ОГЛАВЛЕНИЕ

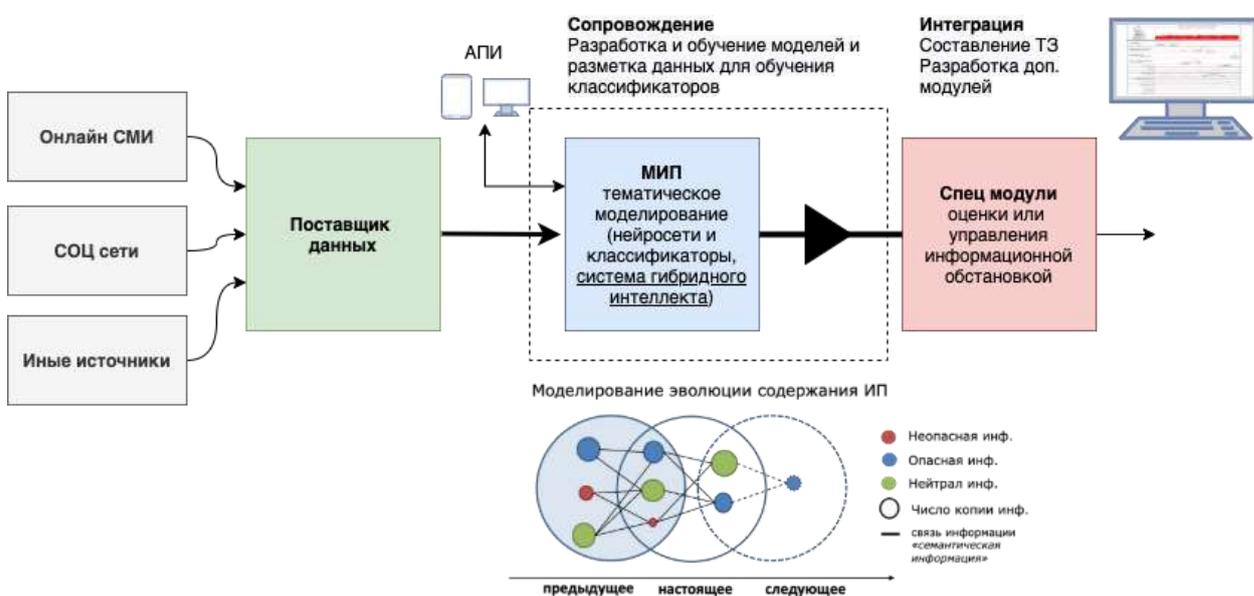
ОБЩАЯ ИНФОРМАЦИЯ	3
ВОЗМОЖНОСТИ ПРОГРАММНЫХ МОДУЛЕЙ ПК МИП	5
1. СБОР НЕСТРУКТУРИРОВАННЫХ ДАННЫХ ОТ ПОСТАВЩИКОВ	5
2. СТРУКТУРИЗАЦИЯ И ИНФОРМАЦИОННОЕ ОБОГАЩЕНИЕ ИСХОДНЫХ ДАННЫХ (АПИ)	5
3. МОДЕЛИРОВАНИЕ СОДЕРЖАНИЯ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА	5
4. МОДЕЛИРОВАНИЕ ИЗМЕНЕНИЙ СОДЕРЖАНИЯ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА	7
5. ОЦЕНКА ИЗМЕНЕНИЙ СОДЕРЖАНИЯ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА	8
ВАРИАНТЫ ИСПОЛЬЗОВАНИЯ МИП	9
ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ МИП	10
ИНТЕЛЛЕКТУАЛЬНАЯ СОБСТВЕННОСТЬ	11

Общая информация

МИП – перспективный программный комплекс моделирования информационного пространства (ИП), применяемый для организации процесса мониторинга информационного пространства больших групп людей в режиме опережающего времени.

МИП – решает задачу прогнозирования и оценки изменения содержания информационного пространства по заданному пользователем тематическому направлению.

МИП обеспечивает главное преимущество её пользователю – возможность раннего реагирования на информационные события до того, как они получили массовое распространение.



В основе теоретического базиса МИП лежит математически формализованная модель процесса эволюционного изменения содержания информационного пространства. Процесс эволюции содержания ИП представлен как изменяющееся во времени пространство



тематических событий (смыслов). Программная реализация МИП использует технологии искусственного интеллекта и машинного обучения.

Целевой пользователь МИП – организации, которые осуществляют мониторинг информационного пространства в режиме опережающего времени для решения задач обеспечения информационной безопасности субъектов различного уровня.

Процесс работы МИП подразумевает следующие этапы

1. Сбор неструктурированных данных от поставщиков;
2. Структуризация и информационное обогащение исходных данных;
3. Моделирование содержания информационного пространства;
4. Моделирование изменений содержания информационного пространства;
5. Оценка изменений содержания информационного пространства.

По запросу разрабатываются иные (специальные) серверные и клиентские модули.

Возможности программных модулей ПК МИП

1. Сбор неструктурированных данных от поставщиков

По умолчанию система поставляется с интеграцией к поставщику данных публикаций из Телеграмм.

Разработаны различные механизмы интеграции с ключевыми поставщиками данных публикаций из других социальных медиа и СМИ.

Под заказ выполняется разработка новых «коннектов» для других поставщиков.

2. Структуризация и информационное обогащение исходных данных (АПИ)

С применением технологий ИИ поступающие в систему публикации автоматически обрабатываются на предмет:

- + определения языка;
- + выделения заголовка;
- + определение ключевого содержания;
- + извлечения именованных сущностей (имена, названия, наименование локаций и определение тональности их упоминания (позитивно, нейтрально, негативно));
- + получение семантического вектора (смыслового вектора);
- + выделение ключевых словосочетаний;

3. Моделирование содержания информационного пространства

- + Точная и полная классификация тематик. Обработанные публикации классифицируются по тематикам, заданным пользователем ключевыми словами, а также с помощью методов машинного

обучения, на основе рекомендованных пользователем эталонных публикаций (разметки). Уникальность разработанной гибридной технологии симбиоза пользовательских настроек и машинного обучения позволяет достигать высокой точности классификации – 99% без снижения полноты. На практике это означает что на 1 000 публикаций, некорректно определённых будет не более 10 единиц, что позволяет существенно снизить информационную нагрузку на аналитика данных.

Динамическая кластеризация публикаций в темы. Поток публикаций в течение дня обрабатывается на предмет группировки публикаций в темы по заданным пользователем тематикам, смысловому содержанию и фактам (географии мест, имен персон, наименований организаций, даты и времени событий). Это позволяет выявлять темы, имеющие признаки массового распространения.

+ **Динамическая кластеризация тем в сюжеты.**

По выявленным темам (группы публикаций) строится граф связей публикаций во времени (1-3 недели). На основе графа конструируются наиболее вероятные сюжеты. Это позволяет выявлять закономерности изменения информационного пространства в течение длительного периода.

В результате формируется модель содержания ИП, имеющая четыре уровня объектов, разделяемых по «времени жизни».

Объект модели ИП	Длительность
Тематика (сюжет-тема-публикация)	Период мониторинга
Сюжет (тема во времени)	7 и более дней
Тема (группа публикаций)	24 часа
Публикация тематики	–

4. Моделирование изменений содержания информационного пространства

На основе запатентованного метода (математико-алгоритмической модели) моделируется изменение содержания информационного пространства объектов на пяти уровнях: словосочетания, публикация, тема, сюжет, тематика. По объектам каждого уровня происходит расчет Силы информационного воздействия, отражающую меру воздействия на информационное пространство по тематикам (смыслам) определенных пользователем. Это позволяет прогнозировать потенциальный охват различных тем и выявлять наиболее популярные темы и тематики в горизонте 7 часов или 7 дней.

В результате формируется прогноз содержания ИП для четырех уровней объектов.

Объект модели ИП	Прогноз	
	Сила инф. воздействия	Охват аудитории
Тематика (сюжет-тема-публикации)	7 дней	–
Сюжет (тема во времени)	7 дней	–
Тема (группа публикаций)	6 часов	+
Мем-граммы (словосочетания)	6 часов / 7 дней	+

5. Оценка изменений содержания информационного пространства

Система позволяет автоматически оценивать темы и сюжеты по заданным пользовательским индикаторам: тональности упоминания целевых субъектов или корреляции событий. Это позволяет оперативно оценить, насколько текущая информационная повестка благоприятна, а также – выявлять скрытые закономерности информационных кампаний. В результате формируется прогноз содержания ИП для четырех уровней объектов.

Объект модели ИП	Индикаторы
Тематика (сюжет-тема-публикации)	Сюжеты и темы
Сюжет (тема во времени)	Корреляция сюжетов
Тема (группа публикации)	Тональность упоминания целевых субъектов
Мем-граммы (словосочетания в общ. сознании)	Сила инф. воздействия

Варианты использования МИП

Вариант А. Приобретение доступа к АПИ

Предоставляется настройка и онлайн доступ к модулям МИП:

1. Сбор неструктурированных данных от поставщиков;
2. Структуризация и информационное обогащение исходных данных.

Стоимость определяется по согласованию объема запросов в месяц.

Вариант Б. Интеграция МИП в ИТ-инфраструктуру заказчика

Предоставляется настройка и онлайн доступ к модулям МИП:

1. Сбор неструктурированных данных от поставщиков;
2. Структуризация и информационное обогащение исходных данных;
3. Моделирование содержания информационного пространства;
4. Моделирование изменений содержания информационного пространства;
5. Оценка изменений содержания информационного пространства.

Стоимость определяется после разработки ТЗ.



Технические характеристики МИП

Размер моделируемого ИП:

до 1 млн публикации в сутки на 1 пользователя.

До 50 тематик на одного пользователя

Базовое число публикаций: 250 тыс в сутки.

Типы моделируемого содержания: числовые и именованные сущности, мем-граммы, тематические кластеры

Поддерживаемые языки:

Гарантия хорошего качества: русский, английский, украинский, немецкий, испанский, турецкий, французский

Возможны лингвистические коллизии: 50+ языков мира

Типы обрабатываемых публикации: статьи СМИ, посты соц. сетей

Интеллектуальная собственность

1. Патент на изобретение «Автоматизированная система моделирования изменения содержания информационного пространства» №2717903, в государственном реестре изобретений с 26.03.2020г, приоритет от 14.04.2019
2. Программа для ЭВМ «Автоматизированная система классификации содержания информационного пространства» №2019661748 от 06.09.2019г.
3. Программа для ЭВМ «МИП - система мониторинга содержания информационного пространства в опережающем времени» №2020612285 от 19.02.2020г.