

*Д. А. Михин, Н. С. Толстошеин**

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ СЕМАНТИЧЕСКОГО АНАЛИЗА В ЗАДАЧАХ МОНИТОРИНГА СОСТОЯНИЯ И МОДЕРНИЗАЦИИ ОРГАНИЗАЦИОННО-ТЕХНИЧЕСКИХ СИСТЕМ

Организационно-технические системы (ОТС) являются сложными комплексами, функционирование и эффективность которых напрямую зависят от состояния технических компонентов и их своевременной модернизации. В связи с этим эффективное управление ОТС требует постоянного мониторинга состояния элементов и оперативного принятия решений о необходимости их обновления или замены, что в современном мире все чаще возлагается на информационные системы.

Традиционные подходы, ядром которых является сбор данных с сенсоров, и анализ эксплуатационных отчетов часто оказываются недостаточными, в связи с тем, что не позволяют учитывать при работе большие объемы инновационной информации, в частности: научные исследования, статьи из авторитетных источников, новые подходы и т.д.

В этих условиях перспективным решением является внедрение в ОТС модуля семантического анализа текстовой информации, который позволит извлекать смысловую информацию из документов, выявлять ключевые признаки необходимости модернизации элементов или их состава в рамках ОТС.

Методы семантического анализа позволяют обрабатывать как структурированные, так и неструктурированные данные [1], выявлять ключевые термины, тематические связи и скрытые паттерны, которые невозможно обнаружить вручную. Основные методы автоматизированного семантического анализа представлены на рис. 1.



Рис 1. Основные методы семантического анализа

* Работа выполнена под руководством доктора технических наук, профессора кафедры «Информационные системы и защита информации» ФГБОУ ВО «ПГТУ» В. В. Алексеева.

Методы векторного представления (wordembeddings) – данный вид семантического анализа основан на представлении слов в виде векторов, которые отражают их значение, он помогает находить синонимы и группировать похожие понятия, что особенно полезно при анализе больших массивов текстовой информации. Наиболее известными примерами метода векторного представления служат Word2Vec, GloVe, ELMo и BERT.

Метод TF-IDF (TermFrequency–InverseDocumentFrequency, Частота слова–Обратная частота документа) – позволяет определить наиболее важные термины в документах и выявлять повторяющиеся проблемы, часто используется в задачах семантического анализа текстовых массивов и информационного поиска.

Метод тематического моделирования – позволяет выявить скрытые темы в большом массиве документов, что позволяет видеть, какие компоненты системы чаще всего упоминаются в контексте проблем или модернизации для использования этой информации в рамках задачи принятия решений. Примером метода тематического моделирования является Латентное размещение Дирихле (LDA).

Для наглядного сравнения методов семантического анализа была составлена табл. 1.

1. Сравнение основных методов семантического анализа

		Векторное представление	Тематическое моделирование
Преимущества	Прост в реализации, быстро работает на больших объемах текстов	Учитывает семантику, позволяет работать с синонимами	Позволяет выявлять тренды и тематические группы
Ограничения	Не учитывает контекст, синонимы и омонимы	Требует больших обучающих корпусов	Модель чувствительна к настройкам, требует предобработки
Применимость к ОТС	Выделение ключевых слов в отчетах и документации	Группировка терминов, выявление скрытых связей между проблемами компонентов	Анализ текстовых отчетов, выявление областей системы с высокой частотой упоминаний проблем

Использование описанных выше методов в рамках задачи мониторинга компонентов организационно-технических систем позволяет выделять ключевые признаки из текстов для связи с конкретными элементами системы, формируя тем самым основу для принятия решений о модернизации или изменении состава ОТС.

Анализ результатов сравнения показывает, что для успешного выполнения задач мониторинга и модернизации организационно-технических систем лучшие результаты будут достигнуты при комбинировании различных методов. TF-IDF хорошо работает для быстрой фильтрации ключевых терминов, метод векторного представления позволяет учитывать синонимы и скрытые связи, а тематическое моделирование выявляет общие темы и паттерны. Их совместное использование повышает точность выявления элементов системы, требующих модернизации или замены, на основе используемых текстовых данных.

Внедрение семантического анализа в систему поддержки принятия решений (СППР) для ОТС может быть реализовано в рамках соответствующего модуля. На вход системы поступают данные о состоянии имеющихся элементов организационно-технической системы из отчетов, регламентов, технической документации, а также массив научных публикаций, специализированных статей и подобных им материалов, касающихся технических аспектов функционирования системы, для корректной и своевременной оценки потенциала возможных изменений компонентов ОТС или их состава.

Алгоритмы обработки текстов на основании семантического анализа выявляют признаки проблем и имеющиеся направления модернизации, которые передаются в модуль принятия решений для сопоставления с конкретными элементами ОТС, определения приоритетов модернизации и формирования рекомендаций. Такой подход автоматизирует работу экспертов, ускоряет обработку информации и повышает обоснованность принимаемых решений. Схема получения рекомендаций по модернизации ОТС представлена на рис. 2.



Рис 2. Схема получения рекомендаций по модернизации ОТС

Перспективы дальнейшего развития описываемого подхода включают дополнительную интеграцию семантического анализа с нейронными сетями и моделями машинного обучения. Нейронные сети могут быть использованы для классификации текстов и выявления скрытых паттернов, что снизит время обработки входных данных, также их использование может оказать благоприятное влияние на оценку потенциального воздействия рекомендуемых изменений на функционирование организационно-технической системы в рамках СППР.

Совмещение технологий нейронных сетей с методами семантического анализа позволяет создавать более адаптивные системы поддержки принятия решений, способные непрерывно анализировать поступающую информацию, выявлять, какие элементы системы требуют модернизации, и предлагать наилучшие варианты их обновления.

Дополнительным перспективным направлением является интеграция методов семантического анализа с технологиями прогнозной аналитики и цифровых двойников ОТС. Такой подход позволяет не только анализировать текущее состояние системы и ее компонентов, но и моделировать потенциальные сценарии изменения состояния при различных вариантах модернизации. В результате СППР сможет не только выявлять элементы, требующие обновления, но и прогнозировать последствия предлагаемых изменений, оценивая риски и ресурсоемкость модернизации.

В заключение отметим, что совмещенное применение методов семантического анализа в задачах мониторинга состояния и модернизации организационно-технических систем демонстрирует высокую эффективность и наличие потенциала для дальнейшего развития. Интеграция семантического анализа в систему поддержки принятия решений (СППР) позволяет обрабатывать не только внутренние отчеты и регламенты, но и внешние научные публикации и статьи, что дает возможность более обоснованно оценивать потенциал изменений и формировать рекомендации на основе передового опыта в рамках системы поддержки принятия решений.

Список литературы

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Е. И. Большакова, Э. С. Клышинский и др. – М. : МИЭМ, 2011. – 272 с.

*Кафедра «Информационные системы и защита информации»
ФГБОУ ВО «ТГТУ»*