ISSN 2658-5014 (Print) СИИТ ISSN 2686-7044 (Online)
2024. Т. 6, № 2 (17). С. 50–65 http://siit.ugatu.su
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
*Научный обзор*

УДК 004.65 DOI 10.54708/2658-5014-SIIT-2024-no2-p50 EDN NLDWBE

# ОЦЕНКА АКТУАЛЬНЫХ УГРОЗ И УЯЗВИМОСТЕЙ ОБЪЕКТОВ КРИТИЧЕСКОЙ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ТЕКСТОВ

**Н. В. Кучкарова**

**Аннотация.** Представлен обзор результатов исследования объектов критической информационной инфраструктуры в плане оценки и анализа актуальных угроз безопасности информации и уязвимостей программного обеспечения с целью повышения достоверности и оперативности на основе открытых баз данных и технологий интеллектуального анализа текстов (Text Mining). Для достижения этой цели в работе решались следующие задачи исследования: анализ современного состояния в области автоматизации процесса оценки и анализа актуальных угроз безопасности информации и уязвимостей программного обеспечения объектов критической информационной инфраструктуры; разработка алгоритмов автоматической классификации и суммаризации текстов, содержащихся в специализированных открытых источниках в области информационной безопасности; разработка метода и алгоритма оценки и приоритизации множества угроз безопасности информации для выявленных уязвимостей промышленных автоматизированных систем управления технологическими процессами с использованием технологии семантического анализа текстов; разработка алгоритма построения графовой модели сценария реализации угроз безопасности информации на основе алгоритмов векторного вложения и технологии трансформеров; разработка архитектуры и программного обеспечения исследовательского прототипа интеллектуальной системы поддержки принятия решений в процессе оценки угроз безопасности информации и уязвимостей программного обеспечения объектов критической информационной инфраструктуры; исследование эффективности ее применения при решении практических прикладных задач.

**Ключевые слова:** уязвимости программного обеспечения; угрозы информационной безопасности; Text Mining; векторное представление текстов; модели-трансформеры; семантическая близость; система поддержки принятия решений.

## Введение

Масштабная цифровизация различных сфер экономики, связанная с активным развитием информационных технологий, и переход на удаленный формат работы, вызванный пандемией коронавирусной инфекции в 2019–2021 гг., спровоцировали резкий рост активности киберпреступников. Чаще всего компьютерным атакам подвергаются государственные и медицинские учреждения, промышленные предприятия. Предприятия упомянутых отраслей, как правило, относятся к субъектам критической информационной инфраструктуры (КИИ), являющихся собственниками различных классов объектов КИИ, большую группу которых составляют промышленные автоматизированные системы управления технологическими процессами. Обеспечение безопасности объектов КИИ является, в соответствии с Доктриной информационной безопасности Российской Федерации, одним из приоритетных направлений в области информационной безопасности (ИБ). Требования к обеспечению ИБ объектов КИИ закреплены в ряде нормативно-правовых документов, принятых в России в последние годы, таких как: Федеральный закон «О безопасности критической информационной инфраструктуры» № 187-ФЗ (2017), Приказы ФСТЭК России №№ 31, 235 и 239 (2017), «Методика оценки угроз безопасности информации» ФСТЭК России от 5 февраля 2021 г. Согласно данной методике, одними из ключевых этапов оценки угроз БИ для объектов КИИ являются оценка возможности реализации угроз БИ и определение их актуальности. Реализация данного этапа

связана с необходимостью определения источника угроз, то есть актуальных нарушителей, а также построения сценариев атак. Определение сценариев атак предусматривает, в свою очередь, установление последовательности возможных тактик и соответствующих им техник, применение которых возможно актуальным нарушителем с соответствующим уровнем возможностей посредством эксплуатации уязвимостей программного обеспечения (ПО).

Вместе с тем на практике при решении данной задачи специалисты по ИБ сталкиваются с необходимостью обращения к большим массивам текстовых описаний компонентов (угроз БИ, уязвимостей ПО, тактик, техник), размещенных в открытых базах данных (БД). Так, на момент проведения исследования только Банк данных угроз безопасности информации (БДУ) ФСТЭК России содержал текстовые описания более 220 угроз безопасности информации (БИ), свыше 36 500 уязвимостей ПО, 10 тактик и от 7 до 29 соответствующих им техник. Работа с указанными БД предполагает поиск и анализ угроз БИ и уязвимостей ПО в «ручном» режиме, что требует больших временных затрат и сопровождается ошибками обработки, обусловленными человеческим фактором, в связи с чем закономерно желание специалистов по ИБ автоматизировать процесс сопоставления угроз БИ, уязвимостей ПО, тактик и техник их эксплуатации. Однако существующие на данный момент отдельные разработки в области автоматизации процесса оценки угроз БИ не учитывают в полной мере требований методики относительно определения сценариев атак. В связи с этим тема статьи, посвященная вопросам автоматизации оценки и приоритизации актуальных угроз БИ, уязвимостей ПО, тактик (техник) и построения сценариев возможных атак с использованием методов интеллектуального анализа текстов, является актуальной.

## Степень разработанности темы и постановка задачи

В настоящее время в данной предметной области ведутся активные исследования, о чем свидетельствуют работы ряда отечественных и зарубежных ученых: И. В. Аникина, Д. В. Бондарчука, И. П. Болодуриной, В. И. Васильева, А. М. Вульфина, Р. В. Жука, П. Д. Зегжды, А. С. Катасёва, И. В. Котенко, А. М. Лаврентьева, А. С. Макаряна, И. В. Машкиной, А. Г. Остапенко, В. И. Петренко, А. И. Рагозина, Д. М. Рябова, В. В. Селифанова, А. А. Сычугова, Ф. Б. Тебуевой, F. Alfaycz, M. H. de Boer, E. Hemberg, Y. Lee, Z. Q. Liu, O. Mendsaikhan, S. Noel, S. Shin, V. Smyth, H. Xiao, J. Y. Zhang и др.

Анализ результатов проведенных исследований показал, что при всей их значимости проблема оценки и анализа актуальных угроз БИ и уязвимостей ПО объектов КИИ нуждается в дальнейшей проработке. Существующие подходы ориентированы в основном на установление связей между выявленными уязвимостями ПО и соответствующими им угрозами БИ, оценке степени опасности этих уязвимостей, необходимости использования для этих целей различных источников информации, и в первую очередь, открытых текстовых БД, регулярно пополняемых новыми данными об угрозах и уязвимостях. Вместе с тем сегодня остаются открытыми вопросы комплексной оценки угроз БИ, уязвимостей ПО, тактик и техник их использования, построения сценариев реализации атак на объекты КИИ, а также задачи автоматизации соответствующих процедур, решение которых позволило бы значительно снизить трудоемкость обработки текстовых данных, используемых на этом этапе, повысить достоверность и оперативность принимаемых решений в процессе оценки рисков ИБ и уровня защищенности объектов КИИ.

Объектом данного исследования являются объекты критической информационной инфраструктуры. Предмет исследования – методы и алгоритмы оценки и анализа актуальных угроз безопасности информации и уязвимостей ПО объектов КИИ. Целью является повышение достоверности и оперативности оценки актуальных угроз БИ и уязвимостей ПО объектов КИИ на основе открытых баз данных и технологий интеллектуального анализа текстов. Для достижения этой цели решались следующие задачи:

1. Разработка алгоритмов автоматической классификации и суммаризации текстов, содержащихся в специализированных открытых источниках в области информационной безопасности.

2. Разработка метода и алгоритма оценки и приоритизации множества угроз БИ для выявленных уязвимостей промышленных автоматизированных систем управления технологическими процессами (ПО АСУ ТП) с использованием технологии семантического анализа текстов.

3. Разработка алгоритма построения графовой модели сценария реализации угроз БИ на основе алгоритмов векторного вложения и технологии трансформеров.

4. Разработка архитектуры и ПО исследовательского прототипа интеллектуальной системы поддержки принятия решений в процессе оценки угроз БИ и уязвимостей ПО объектов КИИ, исследование эффективности ее применения при решении практических прикладных задач.

Для решения поставленных задач были использованы методы интеллектуального анализа данных и защиты информации, методы экспертных оценок, теории искусственных нейронных сетей, методология функционального моделирования (IDEF0), методы объектно-ориентированного анализа и проектирования, модели теории графов, методы имитационного моделирования.

Для решения задач проведен анализ современного состояния в области обеспечения ИБ объектов КИИ, дается краткий обзор открытых баз данных (NVD, CVE, CAPEC, БДУ ФСТЭК России и др.), содержащих текстовые описания угроз ИБ, уязвимостей ПО, тактик (техник) их реализации [1]. Проведен анализ основных требований и этапов методики оценки угроз БИ ФСТЭК России. Показано, что в связи с возрастающим объемом слабоструктурированных текстовых описаний угроз БИ и уязвимостей ПО, содержащихся в различных БД, для их обработки и анализа целесообразно использование современных технологий интеллектуального анализа текстов (Text Mining) [2–6].

Проведен сравнительный анализ известных методов и алгоритмов обработки текстов на естественном языке (ЕЯ). В качестве базовых алгоритмов для решения этой задачи выделены: Word2Vec и Doc2Vec – алгоритмы векторного вложения (Word Embedding), позволяющие получить векторное представление слов и предложений (абзацев, документов); TF-IDF – статистическая мера, используемая для оценки важности слова в контексте документа, используется при расчёте меры близости документов при их классификации (кластеризации); BERT – нейросетевая языковая модель, основанная на архитектуре трансформера, предназначенная для предобучения и обработки больших корпусов текстов на ЕЯ. Рассмотрена практика использования данных технологий и алгоритмов в различных прикладных областях, в том числе в области ИБ. Проведенный анализ показал перспективность использования технологий интеллектуального анализа текстов (Text Mining) для решения сформулированных задач исследования.

## Возможность применения технологий Text Mining

Исследована возможность применения технологий Text Mining при решении задач автоматической классификации (тематического моделирования) и суммаризации текстов из открытых источников в области ИБ. Рассмотрены общие подходы к решению задач автоматической классификации и суммаризации текстов. Приведены результаты экспериментов по автоматической классификации слабоструктурированой информации на примере корпуса текстов, сформированного из 438 полнотекстовых научных статей, опубликованных в журнале «Вопросы кибербезопасности» за 2013–2022 гг. [7–10]. В процессе этих экспериментов исследовались различные подходы к классификации (кластеризации) указанных документов: с предварительным понижением размерности признакового пространства и использованием метода ближайших соседей (K-Means); с помощью скрытого распределения Дирихле (LDA) и неотрицательной матричной факторизации (NMF); на основе моделей векторных вложений

(Text Rank, SBert). Были выполнены также эксперименты по автоматической суммаризации текстов для формирования кратких рефератов, раскрывающих смысловое содержание документа. Использование данных технологий позволяет повысить качество анализа текстовых документов в области ИБ и одновременно снизить когнитивную нагрузку на эксперта.

## Функциональная модель процесса оценки и анализа актуальных угроз

Разработана функциональная модель процесса оценки и анализа актуальных угроз БИ в соответствии с методикой ФСТЭК. Показано, что основные сложности при работе с данной методикой связаны с определением сценариев реализации угроз БИ, в силу необходимости использования при этом разнородной слабоструктурированной информации, включающей перечни актуальных уязвимостей ПО, типы доступа к информационным активам, типы нарушителей и т. п., а также ввиду отсутствия эффективных средств автоматизации, позволяющих формализовать и упростить процесс сопоставления имеющихся исходных данных (актуальных уязвимостей ПО, тактик (техник) их реализации, возможных угроз БИ) [11, 12].

Разработаны метод и алгоритм решения задачи автоматизации оценки и приоритизации актуальных угроз БИ и уязвимостей ПО объектов КИИ с использованием технологий интеллектуального анализа текстов (Text Mining). Проанализированы ключевые этапы обработки текстовых описаний с применением указанных технологий, рассмотрены особенности использования алгоритмов Word2Vec, Doc2Vec, BERT в задачах обработки текстовых описаний угроз БИ и уязвимостей ПО [13].

Посредством агрегации данных, содержащихся в БДУ ФСТЭК России, сформирован корпус русскоязычных текстов – описаний угроз БИ, уязвимостей ПО, тактик и техник их реализации. Для анализа данного корпуса разработан алгоритм векторного представления текстовых описаний угроз БИ, уязвимостей ПО, тактик (техник) и оценки семантической близости (сходства) этих описаний. В ходе выполнения предложенного алгоритма загружаются данные из БДУ ФСТЭК, затем они нормализуются, производится экспертная разметка для выделения семантических особенностей текста, строится модель TF-IDF, позволяющая оценить важность каждого слова в корпусе, далее происходит векторизация текстовых документов с использованием моделей Word2Vec, Doc2Vec, BERT с учетом частоты встречаемости (TF-IDF) слов в корпусе.

Для выявления устойчивой структуры текстовых описаний в пространстве признаков векторных вложений на основе оценки их семантической близости разработан алгоритм кластеризации текстовых описаний угроз БИ, уязвимостей ПО, тактик и техник. Приведены результаты кластеризации текстовых описаний с предварительно заданными центрами кластеров, полученных из векторного представления описаний информационных ресурсов и компонентов промышленных систем и сетей объектов КИИ (22 объекта воздействия, согласно методике ФСТЭК) методом K-Means в признаковом пространстве. Визуализация результатов кластеризации с использованием алгоритма снижения размерности признакового пространства t-SNE (t-distributed Stochastic Neighbor Embedding) свидетельствует о наличии структуры компактных групп текстовых описаний (рис. 1).

Разработан алгоритм оценки и приоритизации актуальных угроз БИ и уязвимостей ПО (рис. 2) с использованием технологий Text Mining. Для апробации предложенного метода и алгоритма анализировались текстовые данные из БДУ ФСТЭК. Из общего объема текстовых описаний в 740634 слова был сформирован словарь, содержащий 12884 слова. После процедуры предобработки и нормализации данных была построена модель векторных вложений Doc2Vec с помощью фреймворка Gensim.

В процессе тестирования работы предложенного алгоритма были рассмотрены несколько уязвимостей ПО, выявленных хостовым сканером уязвимостей, в частности уязвимость BDU:2015-00285 «Уязвимость программного обеспечения Flash Player, позволяющая удаленному злоумышленнику нарушить конфиденциальность, целостность и доступность защищаемой информации». Данной уязвимости эксперт в ручном режиме сопоставил угрозу УБИ.192.

Используя текстовое описание выбранной уязвимости, с помощью разработанного алгоритма был проведен выбор семантически близких по описанию угроз БИ из БДУ ФСТЭК.

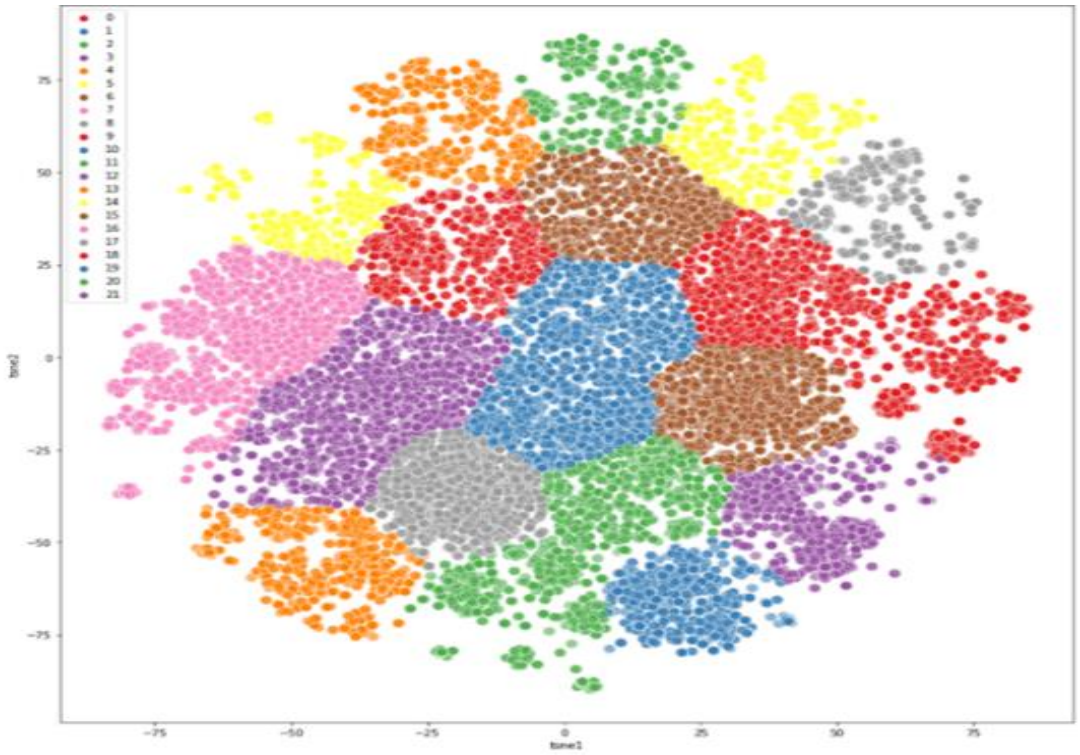

**Рис. 1** T-SNE визуализация кластерной структуры с предварительно заданными центрами.

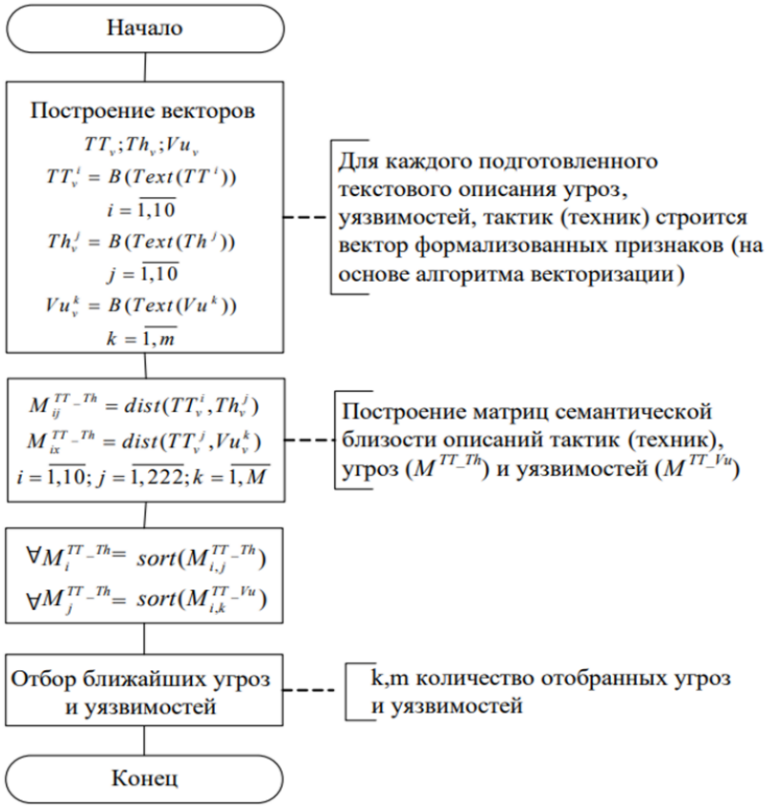


**Рис. 2** Блок-схема алгоритма приоритизации множества угроз БИ для выявленных уязвимостей ПО на основе сопоставления текстовых описаний угроз БИ (Th), уязвимостей (Vu), техник (тактик) (TT).

На рис. 3 показаны результаты подбора 10 релевантных угроз БИ, ранжированных в порядке убывания метрики семантической близости. Как видно, угроза УБИ.192 попадает в данный перечень, что совпадает с результатом предварительного экспертного оценивания, но разработанный алгоритм предлагает расширенный перечень релевантных угроз БИ, которые также должны быть приняты во внимание.


**Рис. 3** Релевантные угрозы БИ, отсортированные в порядке убывания метрики семантической близости (score) к выявленной уязвимости BDU:2015-00285.

Как показывают экспертные оценки, финальная стадия анализа позволяет значительно упростить сопоставление актуальных угроз БИ и уязвимостей ПО для конкретных версий ПО и сократить количество просматриваемых экспертом угроз БИ для каждой отдельной уязвимости ПО более чем в 10–15 раз.

Результаты проведенных исследований показывают возможность построения семантической графовой модели сценария реализации угроз БИ. Общая схема графа соответствия множеств угроз БИ, уязвимостей ПО, тактики техник их эксплуатации показана на рис. 4.

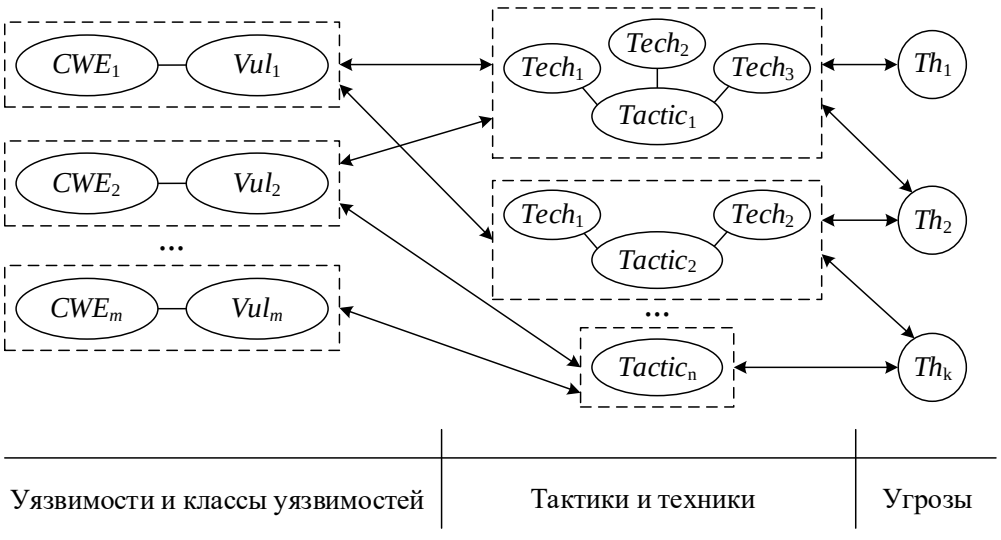


**Рис. 4** Граф соответствия множеств угроз БИ, уязвимостей ПО, тактик и техник.

Блок-схема алгоритма построения графовой модели фрагмента сценария реализации угроз БИ (графа соответствия) представлена на рис. 5. На вход алгоритма подаются текстовые данные из подготовленного корпуса текстов, далее происходят выбор модели векторного представления текстов, их предобработка и векторизация. Затем в несколько этапов строится указанный граф соответствия, то есть происходит соотнесение множеств выявленных уязвимостей и слабостей ПО, затем – уязвимостей и слабостей ПО с тактиками и техниками и в заключение – тактики и техники соотносятся с угрозами БИ на основе оценки семантической близости их описаний. На выходе алгоритма эксперты проводят оценку качества полученной графовой модели.

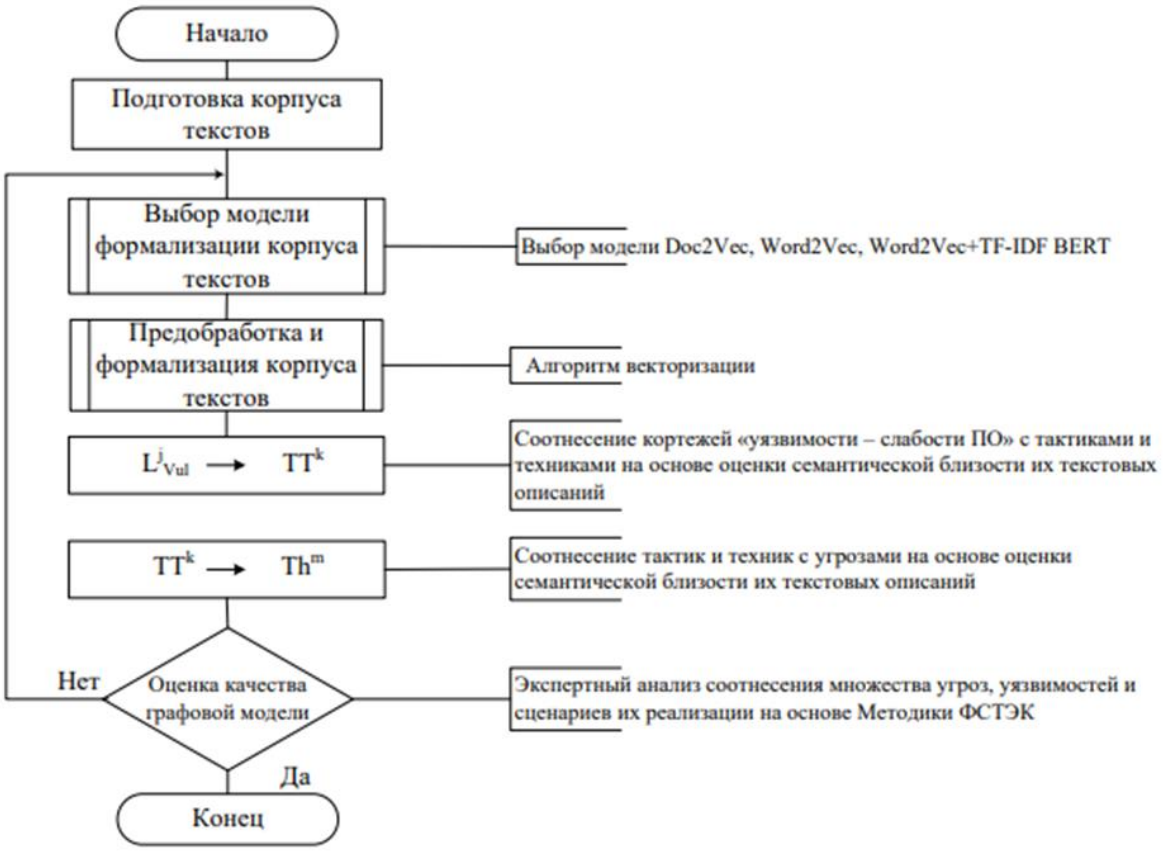


**Рис. 5** Блок-схема алгоритма построения графа соответствия множеств угроз БИ, уязвимостей ПО, тактик и техник их эксплуатации.

В основе алгоритма оценки и приоритизации множества угроз БИ для выявленных уязвимостей ПО лежит построение матриц оценок попарной семантической близости элементов двух множеств: $M^{TT\_Vu}$ тактик (техник) (TT) и уязвимостей (Vul), $M^{TT\_Th}$ тактик (техник) (TT) и угроз (Th). Далее матрицы сортируются построчно в порядке убывания метрики семантической близости текстовых описаний элементов множеств с обрезкой по количеству элементов в строке: для угроз БИ остается $p = 10$ наиболее схожих, для уязвимостей ПО $q = 25$ наиболее схожих с текстовым описанием тактик (техник). В качестве меры оценки семантической близости используется косинусная мера, позволяющая вычислить расстояние между двумя векторами в семантическом пространстве признаков:

$$similarity = cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{n=1}^{i} A_i B_i}{\sqrt{\sum_{n=1}^{i} A_i^2} \cdot \sqrt{\sum_{n=1}^{i} B_i^2}}, \quad (1)$$

где $A$ *и* $B$ – векторные представления текстовых описаний сопоставляемых пар элементов: уязвимостей ПО, тактик, техник и угроз БИ.

Общий порядок моделирования и оценки актуальности угроз БИ на основе перечня актуальных уязвимостей ПО определен в методике ФСТЭК. В соответствии с данной методикой, предполагаются стратегический и тактический уровни построения модели угроз БИ. Стратегический уровень включает в себя определение типа нарушителя, цели воздействия, негативные последствия, а тактический – применяемые тактики и техники эксплуатации уязвимостей ПО (то есть возможные сценарии реализации угроз БИ). Семантическая модель сценария реализации угроз БИ (рис. 6) представляет собой граф

$$G = \{V, E, D\}, \quad (2)$$

где $V$ – множество вершин графа – текстовые описания угроз БИ, уязвимостей ПО, тактик и техник:

$$V = V_1 \cup V_2 \cup V_3 \cup V_4; \quad (3)$$

$V_1$ – множество вершин, соответствующих идентификаторам выявленных уязвимостей ПО;

$V_2$ – множество вершин, соответствующих техникам реализации атаки, которые описывают инструменты, технологии, утилиты и т. д., используемые нарушителем;

$V_3$ – множество вершин, соответствующих тактикам, то есть действиям на разных этапах реализации атаки;

$V_4$ – множество вершин, соответствующих актуальным угрозам БИ;

$E$ – множество взвешенных ориентированных ребер, устанавливающих отношения между текстовыми описаниями:

$$E \subseteq V \times V, \;\; e(v_i, v_j), \;\; v_i, v_j \in V; \quad (4)$$

$D(e)$ – функция, определяющая степень семантической близости для концептов $v_i, v_j \in w$;

$w$ = dist $(v_i, v_j)$ – весовой коэффициент, характеризующий метрику семантической близости текстовых описаний смежных вершин.

Семантическая модель фрагмента сценариев реализации угроз $G$ объектов АСУ ТП строится на основе перекрестных ссылок в гипертекстовых документах (текстовых описаний угроз БИ, уязвимостей ПО и тактик (техник) их реализации – $V$ графа). Ребра модели нагружаются весовыми коэффициентами $w^{i,j}$, характеризующими значения метрики семантической близости текстовых описаний смежных вершин, полученные с помощью методов машинного обучения для векторного представления текстов документов (Word2Vec, Doc2Vec и технологии трансформера).

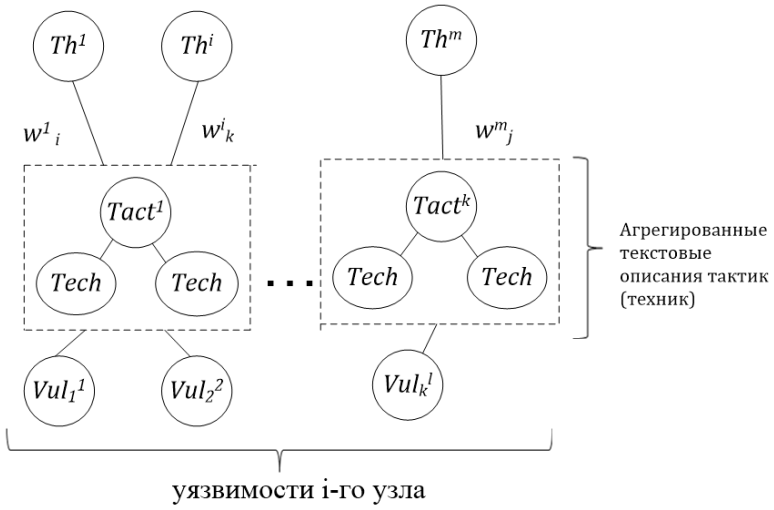


**Рис. 6** Графовая модель фрагмента сценария реализации угроз.

Алгоритм построения графовой модели фрагмента сценария реализации угроз с использованием Word2Vec, Doc2Vec и технологии трансформера представлен на рис. 7. Построение графовой модели начинается с подготовки текстовых описаний угроз, уязвимостей, тактик и техник. Затем на основе ссылочных описаний устанавливаются связи между вершинами $V_1$, $V_2$, $V_3$ , $V_4$ . Далее строится матрица семантической близости описаний угроз БИ, уязвимостей ПО, тактик и техник, несуществующие связи прореживаются на основе порогового значения. В заключение производятся оценка и приоритизация рассмотренных множеств угроз БИ и уязвимостей ПО.

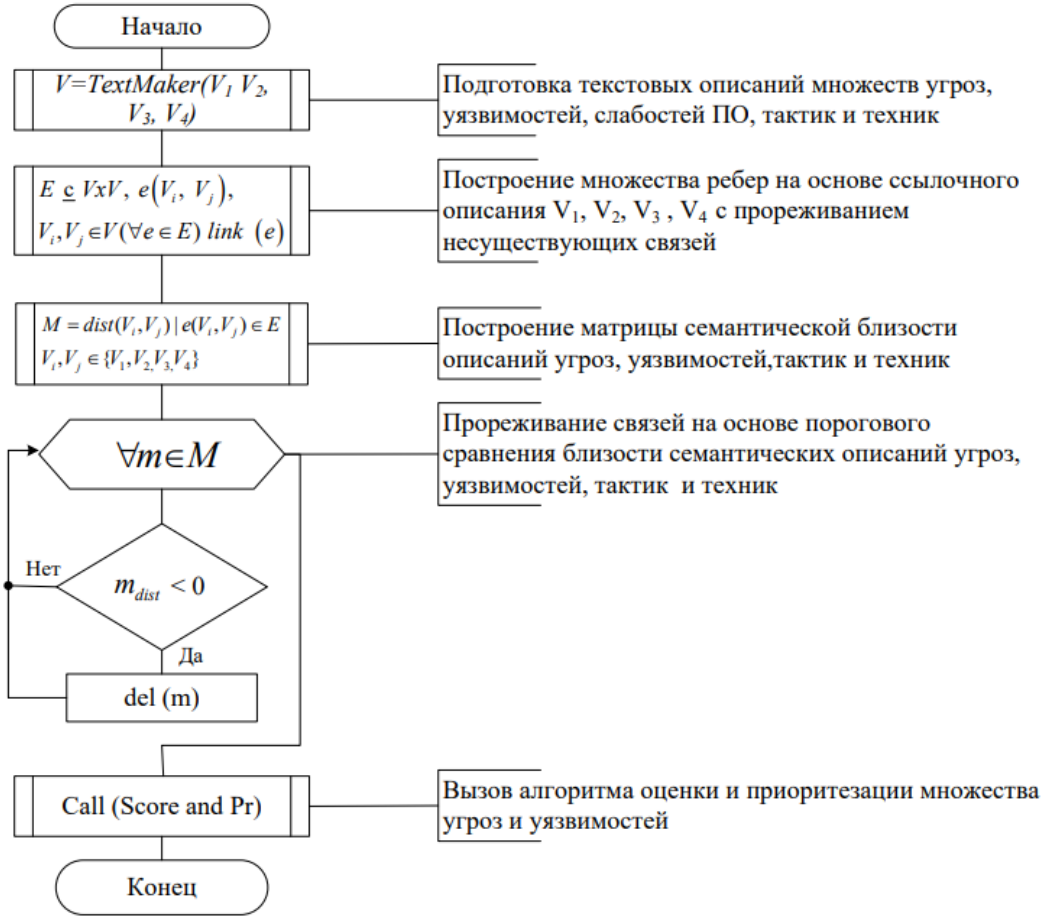


**Рис. 7** Блок-схема алгоритма построения графовой модели фрагмента сценария реализации угроз БИ.

Далее, при построении графовой модели, описывающей отношения множеств угроз БИ, уязвимостей ПО, тактик и техник их эксплуатации, используется технология трансформеров, в частности нейросетевая языковая модель BERT (Bidirectional Encoder Representations from Transformers), основанная на объединении стека нейросетевых кодировщиков с механизмом внутреннего внимания (Self-Attention). Особенностью построения трансформера является двунаправленная обработка входных слов, что сокращает вычислительные затраты и повышает качество обучения языковой модели [14].

Дальнейший анализ (сопоставление) формализованных представлений текстовых описаний угроз БИ, уязвимостей ПО, техник (тактик) основан на применении методов кластеризации многомерных данных. Визуализация полученной графовой модели показывает, что наилучший результат демонстрирует предобученная модель-трансформер BERT – Large Model. Модель ruBERT-tiny (дистиллированная модель-трансформер многозадачного обучения) также демонстрирует заметное распределение на компактные группы объектов в семантическом векторном пространстве в соответствии с семантической близостью их описаний.

Результаты проведения эксперимента с применением модели BERT – Large Model для заданной уязвимости BDU:2021-02033 показывают высокую степень совпадения экспертных оценок и предложенных сценариев реализации угроз БИ.

## Исследовательский прототип ИСППР

Разработан исследовательский прототип ИСППР для автоматизации процесса оценки и анализа актуальных угроз БИ объектов КИИ [12, 13].

Система включает в себя следующие основные компоненты (программные модули):

– подсистему локального хранения актуальной копии данных из БДУ ФСТЭК (I);

– подсистему сопоставления уязвимостей ПО, тактик (техник) и угроз БИ на основе оценки близости их текстовых описаний (II);

– подсистему оценки актуальных угроз БИ и уязвимостей ПО для объекта КИИ (III).

Схема структурно-функциональной организации подсистемы анализа актуальных угроз БИ и уязвимостей ПО представлена на рис. 8.

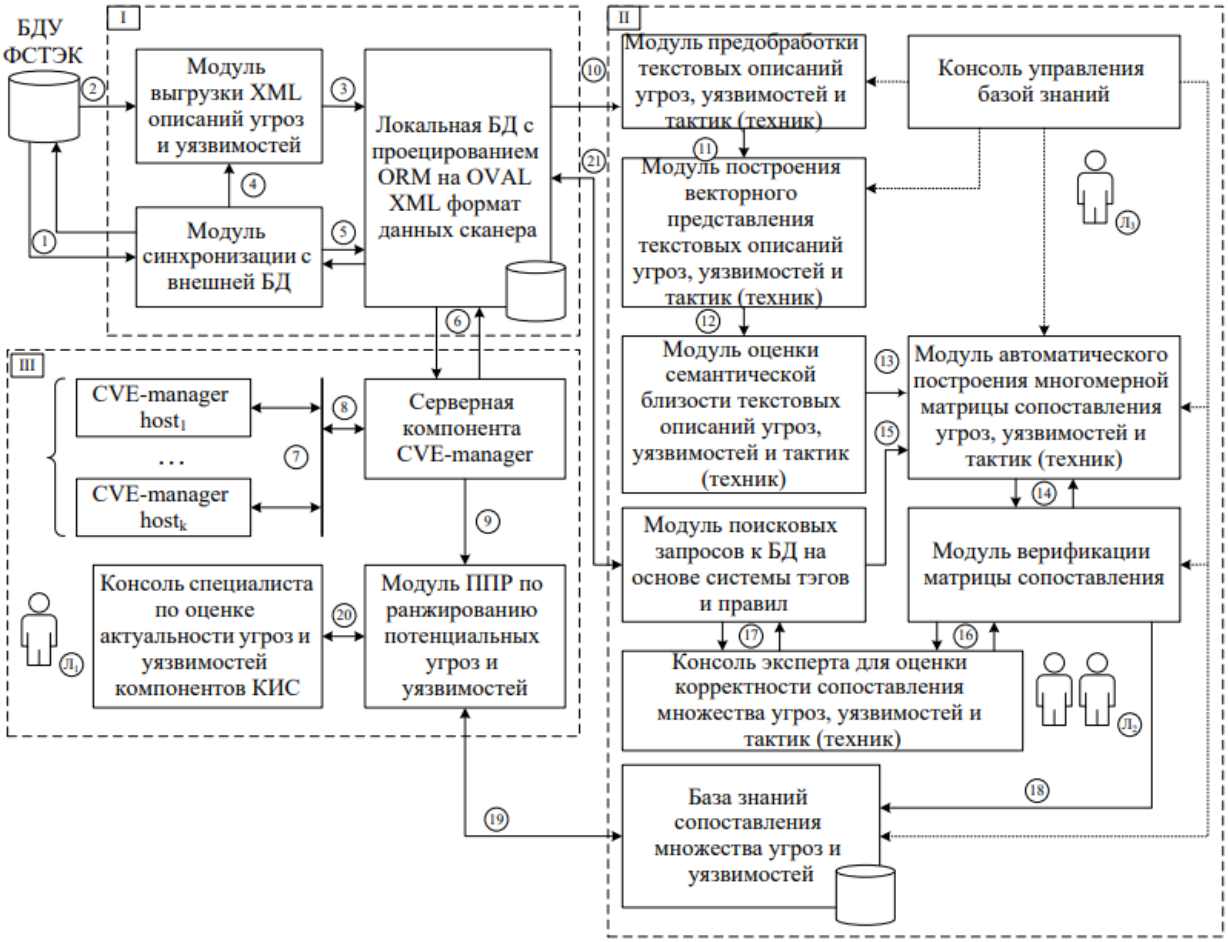


**Рис. 8** Структурно-функциональная организация подсистемы анализа актуальных угроз БИ и уязвимостей ПО

На рис. 9 представлена архитектура программного обеспечения ИСППР, построенная в нотации диаграммы компонент UML.

С целью оценки эффективности применения разработанной ИСППР в работе рассматривалась задача оценки и анализа конкретного промышленного объекта КИИ – АСУ ТП приема, хранения и отпуска товарной нефти. В соответствии с ГОСТ Р 62443 были построены базовая и референсные модели архитектуры данного объекта, а также зональные модели безопасности, выделенные с учетом предъявляемых требований к ИБ.

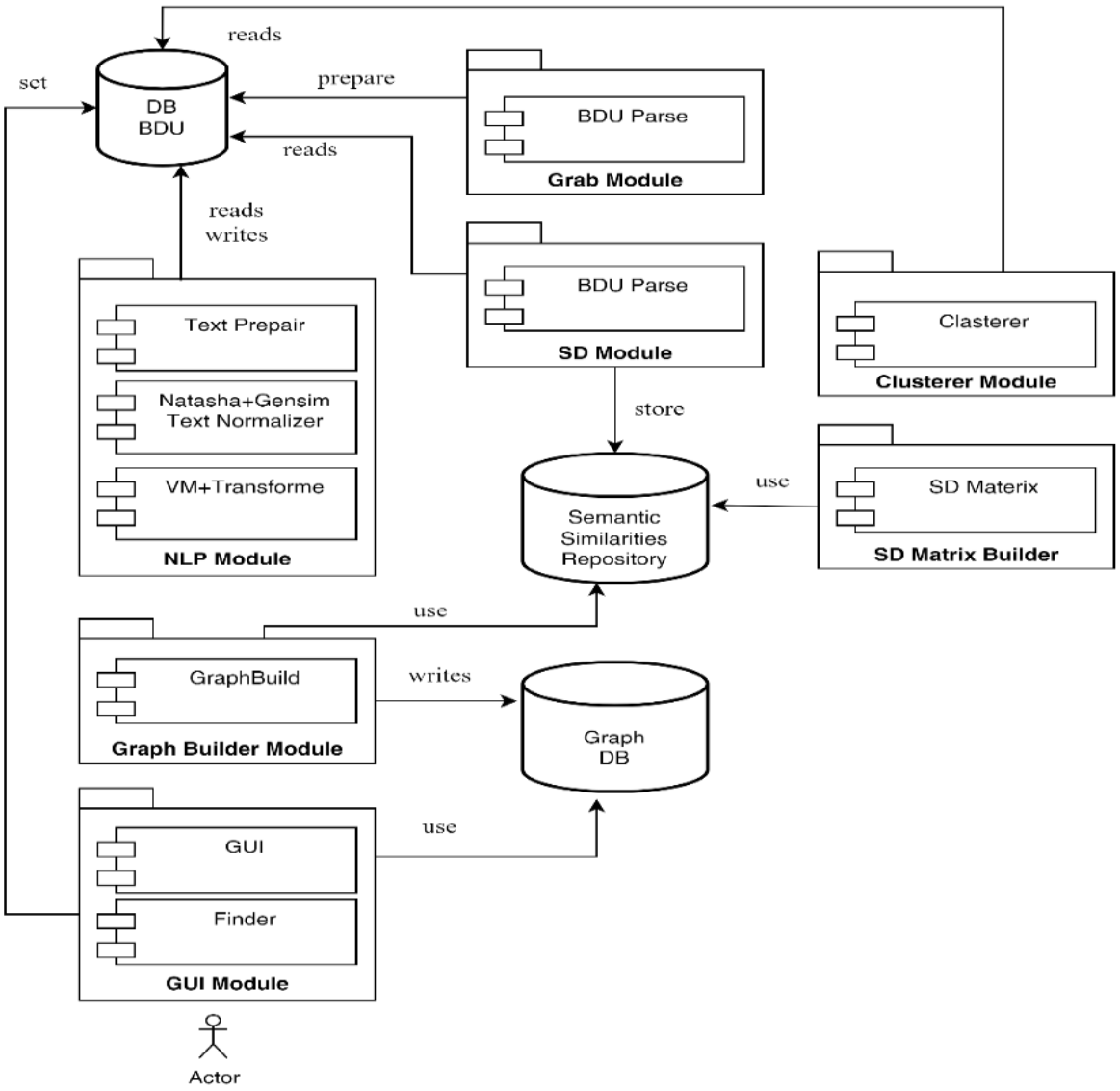


**Рис. 9** Архитектура программного обеспечения ИСППР в нотации UML.

В рамках первой серии экспериментов для анализируемого объекта был сформирован список из 22 наиболее критичных уязвимостей ПО промышленной сети АСУ ТП. В ходе ручного анализа экспертом были выявлены четыре потенциальные угрозы БИ для целевых активов АСУ ТП, время разработки сценариев их реализации составило более 1 часа. Семантический анализ с использованием разработанных технологий интеллектуального анализа текстов позволил автоматизировать процедуру префильтрации множества релевантных угроз БИ и способов их реализации, тем самым существенно сократив затраты времени эксперта.

Результаты экспериментов, проведенных для рассматриваемой АСУ ТП, представлены в виде таблицы сопоставления тактик (техник), наиболее близких угроз БИ и уязвимостей ПО, выявленных с помощью сканеров уязвимостей или определенных экспертом (табл. 1).

Сравнение показателей процедуры анализа уязвимостей ПО, угроз БИ и сценариев их реализации, проведенное экспертом и проведенное с использованием разработанных средств автоматизации, приведено в табл. 2.

Таблица 1

**Фрагмент таблицы сопоставления тактик (техник), угроз БИ и уязвимостей ПО**

| Номер тактики | Текстовое описание тактики и техник | Индексы семантически близких угроз | Текстовое описание семантически близких угроз | Индексы семантически близких уязвимостей | Текстовые описания семантически близких уязвимостей |
|---|---|---|---|---|---|
| 8 | Получение доступа (распространение доступа) к ... | [140, 98, 81, 23, 171, 84, 116, 27, 115, 80] | [Угроза приведения системы в состояние «отказ ... | [BDU:2019-02466, BDU:2019-02818, BDU:2020-0189... | [Уязвимость программного средства централизова... |
| 2 | Получение первоначального доступа к компонентам... | [171, 203, 140, 80, 23, 84, 92, 77, 81, 116] | [Угроза скрытного включения вычислительного устройства... | [BDU:2017-02265, BDU:2017-02264, BDU:2017-0226... | [Уязвимость протокола WPA2, связанная с ошибками... |

Таблица 2

**Сравнение показателей процедуры анализа уязвимостей ПО, угроз БИ и сценариев их реализации**

<table>
<tr><td rowspan="2">Параметр</td><td rowspan="2">Экспертное сопоставление по тегам в БДУ ФСТЭК</td><td colspan="4">Автоматизированная система на основе технологий Text Mining</td></tr>
<tr><td colspan="2">Сопоставление уязвимостей и угроз</td><td colspan="2">Сопоставление уязвимостей, угроз и тактик (техник)</td></tr>
<tr><td>Ввод информации</td><td>Вручную, WEB-интерфейс БДУ</td><td colspan="4">Автоматизированная обработка результатов работы сканеров уязвимостей</td></tr>
<tr><td>Тип сопоставления угроз</td><td>Ручное</td><td colspan="4">Задается пороговыми метриками, определяющими чувствительность фильтра</td></tr>
<tr><td>Количество сопоставленных угроз</td><td>4</td><td colspan="2">10</td><td colspan="2">8</td></tr>
<tr><td rowspan="4">Экспертная оценка корректности сопоставления угроз (техник и тактик)</td><td rowspan="4">—</td><td>Модель</td><td>Оценка</td><td>Модель</td><td>Оценка</td></tr>
<tr><td>Word2Vec + TF-IDF</td><td>6 из 10</td><td>Word2Vec + TF-IDF</td><td>6 из 8</td></tr>
<tr><td rowspan="2">Doc2Vec</td><td rowspan="2">5 и 10</td><td>Doc2Vec</td><td>5 из 8</td></tr>
<tr><td>BERT 3</td><td>7 из 8</td></tr>
<tr><td>Затраченное время на сопоставление угроз и уязвимостей</td><td>Более 15 мин</td><td colspan="2">Менее 5 с</td><td colspan="2">Менее 10 с</td></tr>
<tr><td>Возможность подбора техник и тактик реализации угроз</td><td>Да</td><td colspan="2">Нет</td><td colspan="2">Да</td></tr>
<tr><td>Затраченное время на построение сценариев реализации угроз</td><td>Более 1 часа</td><td colspan="2">—</td><td colspan="2">Менее 20 мин (включая работу эксперта)</td></tr>
</table>

В рамках второй серии экспериментов был использован полный список из 66 выявленных уязвимостей ПО (по группам устройств). Был проведен подбор пороговых значений для оценки семантической близости троек «угроза – уязвимость – тактика (техника)». Результаты экспериментов, проведенных для оценки уязвимостей ПО и угроз БИ рассматриваемой АСУ ТП, приведены в табл. 3 (фрагмент).

Таблица 3

**Результаты сопоставления уязвимостей ПО, угроз БИ, тактик (техник) (фрагмент)**

| | Идентификатор уязвимости | Идентификатор угрозы БИ (УБИ) | № тактики |
|---|---|---|---|
| 3 | BDU:2017-02595 | [11, 21, 25, 53, 57, 64, 73, 75, 95, 106, 107, 110, 116, 119, 122, 131, 134, 142, 150, 173, 194, 210, 215] | [2] |
| 5 | BDU:2018-01029 | | [2, 1] |
| 6 | BDU:2018-01123 | | [2, 5, 1, 9, 6] |
| 10 | BDU:2019-00122 | | [2] |
| 14 | BDU:2019-00516 | | [2] |
| 15 | BDU:2019-00764 | | [2] |
| 17 | BDU:2019-00766 | | [2] |
| 18 | BDU:2019-00767 | | [2] |
| 19 | BDU:2019-01539 | | [2] |
| 21 | BDU:2019-01782 | | [2] |

Для оценки эффективности работы ИСППР был проведен анализ результатов, полученных от ИСППР, на несоответствие мнению эксперта (табл. 4).

Таблица 4

**Оценки несоответствия выданных ИСППР результатов мнению эксперта**

| | | Экспертная оценка | | Сумма |
|---|---|---|---|---|
| | | Положительная | Отрицательная | |
| Оценка ИСППР | Положительная | 11 | 10 | 21 |
| | Отрицательная | 5 | 196 | 201 |
| | сумма | 16 | 206 | 222 |
| | | | | |
| **Accuracy** (точность, правильность измерения) | 0.932 | | | |
| **Precision** (доля объектов, названных классификатором положительными и при этом действительно являющимися положительными**)** | 0.524 | | | |
| **Recall (**доля объектов положительного класса из всех объектов положительного класса, найденные алгоритмом**)** | 0.688 | | | |

Время, затраченное на построение сценариев экспертным способом, составило более 4 часов, при работе эксперта с применением ИСППР затрачено менее 40 мин, согласованность экспертной оценки и ИСППР (F1) составила 0.59.

## Заключение

В рамках исследования получены следующие научные и практические результаты:

1. Разработаны алгоритмы автоматической классификации и суммаризации текстов, содержащихся в специализированных открытых источниках в области ИБ, отличающиеся от известных алгоритмов возможностью осуществлять автоматизированную предобработку больших корпусов слабоструктурированных русскоязычных текстов и их последующий семантический анализ в соответствии с поставленной задачей выделения тематических направлений текстов и их автоматического реферирования.

2. Разработаны метод и алгоритм автоматизированной оценки и приоритизации (ранжирования) множества релевантных угроз БИ для выявленных уязвимостей ПО на основе технологии семантического анализа текстов, отличающиеся использованием предложенного алгоритма кластеризации и оценки семантической близости текстовых описаний угроз БИ и уязвимостей ПО в многомерном векторном пространстве, применение которых позволяет сделать более содержательной работу эксперта, сократив время на поиск актуальных угроз БИ, обеспечивая при этом более высокую наглядность, полноту и достоверность результатов такого поиска.

3. Разработан алгоритм построения графовой модели сценария реализации актуальных угроз БИ на основе оценки семантической близости текстовых описаний соответствующих угроз БИ, уязвимостей ПО, тактик и техник действий нарушителя, отличающийся использованием алгоритмов векторного вложения и технологии трансформеров, что позволяет полностью автоматизировать процесс построения графовой модели, снизить трудоемкость и когнитивную нагрузку на специалистов по ИБ, предоставляя ему дополнительную информацию для формирования перечня актуальных угроз и уязвимостей объектов КИИ.

4. Предложены архитектура и состав программных модулей ИСППР, реализующих предложенные в работе метод и алгоритмы, применение которых позволяет снизить временные затраты и повысить достоверность решений, принимаемых специалистом по ИБ при оценке и анализе актуальных угроз БИ и уязвимостей ПО объектов КИИ.

Результаты представленных разработок внедрены в производственные и бизнес-процессы предприятий г. Уфы: ЗАО «Республиканский центр защиты информации», ООО «Уралтехсистемы», ФГБОУ ВО «Уфимский университет науки и технологий».

Направление дальнейших исследований связано с автоматизацией процесса выбора технических мер и средств обеспечения ИБ объектов КИИ.

## Благодарности и поддержка

Автор выражает признательность научному руководителю проф. В. И. Васильеву, коллегам проф. А. М. Вульфину, проф. В. М. Картаку, проф. И. В. Аникину, А. Д. Кирилловой, которые способствовали выполнению исследования своими консультациями, рекомендациями, критическими замечаниями и научными трудами [12–17].

## Список литературы / References

1. Васильев В. И., Кучкарова Н. В. Подход к определению актуальных уязвимостей при оценке уровня защищенности значимых объектов критической информационной инфраструктуры // Безопасность информационного пространства: Труды XVIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых. Изд-во Магнитогорск. гос. техн. ун-та им. Г. И. Носова, 2019. С. 356–361. [[ Vasiliev V. I., Kuchkarova N. V. “An approach to determining current vulnerabilities when assessing the level of security of significant objects of critical information infrastructure” // Security of the Information Space: Proceedings of the XVIII All-Russian scientific and practical conference of students, graduate students and young scientists. Publishing house Magnitogorsk. state tech. University named after G. I. Nosova, 2019. pp. 356-361. (In Russian). ]]

2. Кучкарова Н. В., Васильев В. И., Вульфин А. М. Сравнительный анализ систем классификаций АСУ ТП объектов критической информационной инфраструктуры // Информационные технологии интеллектуальной поддержки принятия решений (ITIDS'2019): Труды VII Всероссийской научной конференции (с приглашением зарубежных ученых): В 3 т. Уфа, 28–30 мая 2019 года. Т. 2. Уфа: УАТУ, 2019. С. 214–219. EDN IBBPTO. [[ Kuchkarova N. V., Vasilyev V. I., Vulfin A. M. “Comparative analysis of classification systems for automated process control systems of critical information infrastructure objects” // Information Technologies for Intelligent Decision Support (ITIDS'2019): Proceedings of the VII All-Russian Scientific Conference (with the invitation of foreign scientists): in 3 volumes, Ufa, May 28–30, 2019. Volume 2. Ufa: UATU, 2019. pp. 214-219. EDN IBBPTO. (In Russian). ]]

3. Васильев В. И., Кучкарова Н. В., Муслимова К. И. Методика определения актуальных угроз кибербезопасности АСУ ТП на основе стандарта ГОСТ Р 62443 // Сборник избранных статей по материалам научных конференций ГНИИ «Нацразвитие», Санкт-Петербург, 27–31 октября 2018 г. Т. 1. СПб.: ГНИИ «Нацразвитие», 2018. С. 122–126. EDN YRZBGX. [[ Vasiliev V. I., Kuchkarova N. V., Muslimova K. I. “Methodology for determining current threats to the cybersecurity of automated process control systems based on the GOST R 62443 standard” // Collection of selected articles based on scientific conferences of the State Research Institute "National Development", St. Petersburg, 27– October 31, 2018. Volume 1. St. Petersburg: State Research Institute “National Development”, 2018. pp. 122-126. EDN YRZBGX. (In Russian). ]]

4. Кучкарова Н. В. Применение моделей интеллектуального анализа текстов при оценке угроз информационной безопасности // Сборник статей по материалам VI Международной научно-практической конференции «Современные проблемы цивилизации и устойчивого развития в информационном обществе». М.: ИРОК, 2021. С.178–184. [[ Kuchkarova N.V.

"Application of text mining models in assessing threats to information security" // Collection of articles based on the materials of the VI International Scientific and Practical Conference "Modern Problems of Civilization and Sustainable Development in the Information Society". Moscow: IROK, 2021, pp. 178–184. (In Russian). ]]

5. Гузаиров М. Б., Машкина И. В. Управление защитой информации на основе интеллектуальных технологий. М.: Машиностроение, 2013. [[ Guzairov M. B., Mashkina I. V. Information Security Management Based on Intelligent Technologies. Moscow: Mashinostroenie, 2013. (In Russian). ]]

6. Жук Р. В., Дзьобан П. И., Власенко А. В. Определение актуальности угроз информационной безопасности в информационных системах обработки персональных данных с использованием математического аппарата нейронных сетей // Прикаспийский журнал: управление и высокие технологии. 2020. № 1(49). С. 169–178. DOI 10.21672/2074-1707.2020.49.4.169-178. [[ Zhuk R. V., Dzioban P. I., Vlasenko A. V. "Determining the relevance of threats to information security in information systems for processing personal data using the mathematical apparatus of neural networks" // Caspian Journal: Management and High Technologies. 2020. No. 1(49), pp. 169–178. DOI 10.21672/2074-1707.2020.49.4.169-178. (In Russian). ]]

7. Васильев В. И., Вульфин А. М., Кучкарова Н. В. Моделирование и суммаризация текстов в области кибербезопасности // Вопросы кибербезопасности. 2023. № 2(54). С. 2–22. [[ Vasilyev V. I., Vulfin A. M., Kuchkarova N. V. "Modeling and summarization of texts in the field of cybersecurity" // Issues of Cybersecurity. 2023. No. 2(54), pp. 2–22. (In Russian). ]]

8. Кучкарова Н. В. Исследование возможности использования кластерного анализа данных при оценке сценариев реализации угроз безопасности // Мавлютовские чтения: Мат-лы XV Всероссийской молодежной научной конференции: В 7 т. Уфа, 26–28 октября 2021 г. Т. 4. Уфа: УГАТУ, 2021. С. 436–440. EDN WHBAZX. [[ Kuchkarova N.V. "Study of the possibility of using cluster data analysis in assessing scenarios for the implementation of security threats" // Mavlyutov Readings: materials of the XV All-Russian Youth Scientific Conference: in 7 volumes, Ufa, October 26–28, 2021. Volume 4. Ufa: UGATU, 2021. pp. 436-440. EDN WHBAZX. (In Russian). ]]

9. Васильев В. И., Вульфин А. М, Кучкарова Н. В. Использование технологии Text Mining при оценке актуальных угроз и уязвимостей программного обеспечения // Приоритетные направления развития науки и технологий: XXVIII Международная научно-практическая конференция, Тула, 12 марта 2021 г. Тула: Инновационные технологии, 2021. С. 144–149. EDN KMYHUQ. [[ Vasilyev V. I., Vulfin A. M., Kuchkarova N. V. "Using Text Mining technology in assessing current threats and software vulnerabilities" // Priority directions for the development of science and technology: XXVIII International Scientific and Practical Conference, Tula, March 12, 2021. Tula: Innovative Technologies, 2021, pp. 144-149. EDN KMYHUQ. (In Russian). ]]

10. Аралбаев Т. З., Абрамова Т. В., Гетьман М. А. Кластерный анализ как инструмент построения и исследования пространственно-временных моделей угроз // Университетский комплекс как региональный центр образования, науки и культуры: Мат-лы Всероссийской научно-методической конференции (с международным участием). Оренбург, 23-25 января 2020 г. Оренбург: ОГУ, 2020. С. 1401–1405. [[ Aralbaev T. Z., Abramova T. V., Getman M. A. "Cluster analysis as a tool for constructing and studying spatio-temporal threat models" // University Complex as a Regional Center of Education, Science and Culture: Materials of the All-Russian Scientific- methodological conference (with international participation). Orenburg, January 23-25, 2020. Orenburg: OSU, 2020, pp. 1401–1405. (In Russian). ]]

11. Васильев В. И., Вульфин А. М., Кучкарова Н. В. Методика оценки актуальных угроз и уязвимостей на основе технологий когнитивного моделирования и Text Mining // Системы управления, связи и безопасности. 2021. №. 3. С. 110–134. [[ Vasiliev V. I., Vulfin A. M., Kuchkarova N. V. "Methodology for assessing current threats and vulnerabilities based on cognitive modeling and Text Mining technologies" // Control, Communication and Security Systems. 2021, no. 3, pp. 110-134. (In Russian). ]]

12. Васильев В. И., Вульфин А. М, Кучкарова Н. В. Система поддержки принятия решений при оценке актуальных угроз и уязвимостей на основе семантического анализа // Мавлютовские чтения: Статьи XIV Всероссийской молодежной научной конференции, Уфа, 01–03 ноября 2020 г. Т. 5. Ч. 2. Уфа: Уфимский государственный авиационный технический университет, 2020. С. 8. EDN RCBBBS. [[ Vasiliev V. I., Vulfin A. M., Kuchkarova N. V. "Decision support system for assessing current threats and vulnerabilities based on semantic analysis" // Mavlyutov readings: Articles of the XIV All-Russian Youth Scientific Conference, Ufa, 01– November 03, 2020. Volume 5 Part 2. Ufa: Ufa State Aviation Technical University, 2020, p. 8. EDN RCBBBS. (In Russian). ]]

13. Васильев В. И., Вульфин А. М., Кучкарова Н. В. Автоматизация анализа уязвимостей программного обеспечения на основе технологии Text Mining // Вопросы кибербезопасности. 2020. № 4(38). С. 22–31. [[ Vasilyev V. I., Vulfin A. M., Kuchkarova N. V. "Automation of software vulnerability analysis based on Text Mining technology" // Issues of Cybersecurity. 2020. No. 4(38), pp. 22–31. (In Russian). ]]

14. Васильев, В. И., Вульфин А. М., Кучкарова Н. В. Оценка актуальных угроз безопасности информации с помощью технологии трансформеров // Вопросы кибербезопасности. 2022. № 2(48). С. 27–38. [[ Vasiliev, V. I., Vulfin A. M., Kuchkarova N. V. "Assessing current threats to information security using transformer technology" // Cybersecurity Issues. 2022. No. 2(48), pp. 27–38. (In Russian). ]]

15. Бакулин М. А. Управление рисками нарушения информационной безопасности значимых объектов критической информационной инфраструктуры // Системная инженерия и информационные технологии. 2023. Т. 5. № 5(14). С. 78–87. DOI 10.54708/2658-5014-SIIT-2023-no5-p78. EDN CRVUZJ. [[ Bakulin M. A. "Managing the risks of violation of information security of significant objects of critical information infrastructure" // System Engineering and Information Technologies. 2023. Vol. 5, No. 5(14), pp. 78-87. DOI 10.54708/2658-5014-SIIT-2023-no5-p78. EDN CRVUZJ. (In Russian). ]]

16. Вульфин А. М. Модели и методы комплексной оценки рисков безопасности объектов критической информационной инфраструктуры на основе интеллектуального анализа данных // Системная инженерия и информационные технологии. 2023. Т. 5. № 4(13). С. 50–76. DOI 10.54708/2658-5014-SIIT-2023-no3-p50. EDN FJPFKC. [[ Vulfin A. M. "Models and methods for comprehensive assessment of security risks of objects of critical information infrastructure based on intelligent data analysis" // System Engineering and Information Technologies. 2023. Vol. 5, No. 4(13), pp. 50-76. DOI 10.54708/2658-5014-SIIT-2023-no3-p50. EDN FJPFKC. (In Russian). ]]

17. Кириллова А. Д. Оценка рисков информационной безопасности АСУ ТП промышленных объектов методами когнитивного моделирования // Системная инженерия и информационные технологии. 2023. Т. 5. № 4(13). С. 77–93. DOI

10.54708/2658-5014-SIIT-2023-no3-p77. EDN CUEUUP. [[ Kirillova A. D. “Assessing the risks of information security of automated process control systems of industrial objects using cognitive modeling methods” // System Engineering and Information Technologies. 2023. T. Vol, No. 4(13), pp. 77-93. DOI 10.54708/2658-5014-SIIT-2023-no3-p77. EDN CUEUUP. (In Russian). ]]

18. Аникин И. В. Методы и алгоритмы количественной оценки и управления рисками безопасности в корпоративных информационных сетях на основе нечеткой логики // Системная инженерия и информационные технологии. 2023. Т. 5. № 3(12). С. 93–113. DOI 10.54708/2658-5014-SIIT-2023-no3-p93. EDN KUKQGP. [[ Anikin I.V. “Methods and algorithms for quantitative assessment and management of security risks in corporate information networks based on fuzzy logic” // Systems Engineering and Information Technologies. 2023. Vol. 5, No. 3(12), pp. 93-113. DOI 10.54708/2658-5014-SIIT-2023-no3-p93. EDN KUKQGP. (In Russian). ]]

19. Васильев В. И., Картак В. М. Применение методов искусственного интеллекта в задачах защиты информации (по материалам научной школы УГАТУ) // Системная инженерия и информационные технологии. 2020. Т. 2. № 2(4). С. 43–50. EDN ZTQFCW. [[ Vasilyev V. I., Kartak V. M. “Application of artificial intelligence methods in information security problems (based on materials from the scientific school of UGATU)” // System Engineering and Information Technologies. 2020. Vol. 2, No. 2(4), pp. 43-50. EDN ZTQFCW. (In Russian). ]]

*Поступила в редакцию 10 апреля 2024 г.*

**МЕТАДАННЫЕ / METADATA**

**Title:** Assessment of current threats and vulnerabilities of critical information infrastructure objects using text mining technologies.

**Abstract:** An overview of the results of a study of critical information infrastructure (CII) objects is presented in terms of assessment and analysis of current threats to information security and software vulnerabilities to increase reliability and efficiency based on open databases and text mining technologies (Text Mining). To achieve this goal, the following research tasks were solved: analysis of the current state in the field of automation of the process of assessing and analyzing current BI threats and software vulnerabilities of CII objects; development of algorithms for automatic classification and summarization of texts contained in specialized open sources in the field of information security; development of a method and algorithm for assessing and prioritizing a variety of BI threats for identified vulnerabilities in ICS software using semantic text analysis technology; development of an algorithm for constructing a graph model of a scenario for the implementation of BI threats based on vector embedding algorithms and transformer technology; development of architecture and software for a research prototype of an intelligent decision support system (ISDS) in the process of assessing BI threats and software vulnerabilities of CII objects; research into the effectiveness of its use in solving practical applied problems.

**Key words:** software vulnerabilities; information security threats; Text Mining; vector representation of texts; transformers models; semantic proximity; decision support system.

**Язык статьи / Language:** русский / Russian.

**Об авторе / About the author:**

**КУЧКАРОВА Наиля Вакилевна**
ФГБОУ ВО «Уфимский университет науки и технологий», Россия.
Старший преподаватель ин-та информатики, математики и робототехники. Дипл. магистр в области информатики и вычислительной техники (Уфимск. гос. авиац. техн. ун-т, 2020). Канд. техн. наук по методам и системам защиты информации, информационной безопасности (Уфимск. ун-т науки и технологий, 2023).
E-mail: nailya_kuchkarov@mail.ru

**KUCHKAROVA Nailya Vakilevna**
Ufa State Aviation Technical University (UGATU), Russia.
Senior lecturer at the Institute of Computer Science, Mathematics and Robotics. Dipl. Master of Science in Informatics and Computer Engineering (Ufa State Aviation Technical University, 2020). Cand. Tech. Sciences on methods and systems of information protection, information security (Ufa University of Science and Technology, 2023).
E-mail: nailya_kuchkarov@mail.ru