ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS

УДК 004.046

DOI: 10.21822/2073-6185-2024-51-4-164-170 Оригинальная статья /Original article

Методы семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак

В.О. Шабля, С.А. Коноваленко, Е.О. Орлов
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознаменное училище имени генерала армии С.М.Штеменко,
350063, г. Краснодар, ул. Красина, д. 4, Россия

Резюме. Цель. Целью исследования является определение наиболее эффективного метода семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак. **Метод.** Исследование проведено на основании методов семантического анализа состояния процесса функционирования СОПКА. **Результат.** Предложена структурная модель системы семантического анализа состояния процесса функционирования СОПКА, способная в полной мере обеспечить анализ состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак. Определено, что наиболее эффективным методом решения задачи семантического анализа состояния процесса функционирования СОПКА является машинное обучение с применением онтологического моделирования объекта анализа. **Вывод.** Необходимы дальнейшие исследования в части разработки предлагаемой системы семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак.

Ключевые слова: система обнаружения, предупреждения, ликвидации последствий компьютерных атак, семантический анализ, семиотика, онтология, машинное обучение

Для цитирования: В.О. Шабля, С.А. Коноваленко, Е.О. Орлов. Методы семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак. Вестник Дагестанского государственного технического университета. Технические науки. 2024; 51(4):164-170. DOI:10.21822/2073-6185-2024-51-4-164-170

Methods for semantic analysis of the state of the process of functioning of a system for detecting, preventing and eliminating the consequences of computer attacks

V.O. Shablya, S.A. Konovalenko, E.O. Orlov
General S.M. Shtemenko Krasnodar Higher Military Orders of Zhukov
and the October Revolution Red Banner School,
4 Krasina St., Krasnodar 350063, Russia

Abstract. Objective. The aim of the study is to determine the most effective method of semantic analysis of the state of the process of detection, prevention and elimination of consequences of computer attacks. **Method.** The study was conducted based on the methods of semantic analysis of the state of the process of SOPKA operation. **Result.** A structural model of the system of semantic analysis of the state of the process of SOPKA operation is proposed, which is capable of providing an analysis of the state of the process of detection, prevention and elimination of consequences of computer attacks. The most effective method for solving the problem of semantic analysis of the state of the process of SOPKA operation is machine learning using ontological modeling. **Conclusion.** The results indicate the need for further research of the sys-

tem of semantic analysis of the state of the process of detection, prevention and elimination of consequences of computer attacks.

Keywords: system for detecting, preventing and eliminating the consequences of computer attacks, semantic analysis, semiotics, ontology, machine learning

For citation: V.O. Shablya, S.A. Konovalenko, E.O. Orlov. Methods for semantic analysis of the state of the process of functioning of a system for detecting, preventing and eliminating the consequences of computer attacks. Herald of Daghestan State Technical University. Technical Sciences. 2024; 51(4):164-170. DOI:10.21822/2073-6185-2024-51-4-164-170.

Введение. Современная практика применения системы обнаружения, предупреждения и ликвидации последствий компьютерных атак (СОПКА) свидетельствует о необходимости систематического проведения анализа состояния процесса функционирования (ПФ) её структурных элементов [1-2]. При этом существующие подходы к анализу в заданной предметной области основываются исключительно на прямом методе анализа, зависящем от полноты предоставляемых параметрических данных (ПД) о текущем состоянии ПФ СОПКА и реализуемом специалистом по информационной безопасности (ИБ), что в общем негативно влияет на эффективность решения рассматриваемой задачи [3]. Обзор существующих теоретических и практических аспектов в заданной предметной области показал, что возможным решением указанной проблемы является применение семиотического подхода к анализу состояния ПФ СОПКА, в рамках которого реализуются синтаксический, семантический и прагматический этапы анализа.

Постановка задачи. В данной статье рассматривается семантический анализ (СмА), который в настоящее время реализуется в условиях отсутствия единого формализованного подхода, а также базируется на конкретных методах, обладающих определенными недостатками, не позволяющими оперативно решить задачу анализа состояния ПФ СОПКА с учетом особенностей её построения и режимов функционирования (РжФ) [1].

Целью исследования является определение наиболее эффективного метода семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак.

Методы исследования. Рассмотрим основные методы, способные выполнять СмА состояния ПФ СОПКА (рис. 1): метод распознавания именованных сущностей; метод основанный на использовании тезауруса; метод лексического анализа; метод машинного обучения; метод глубокого обучения; метод онтологического моделирования; метод семантического дифференциала.

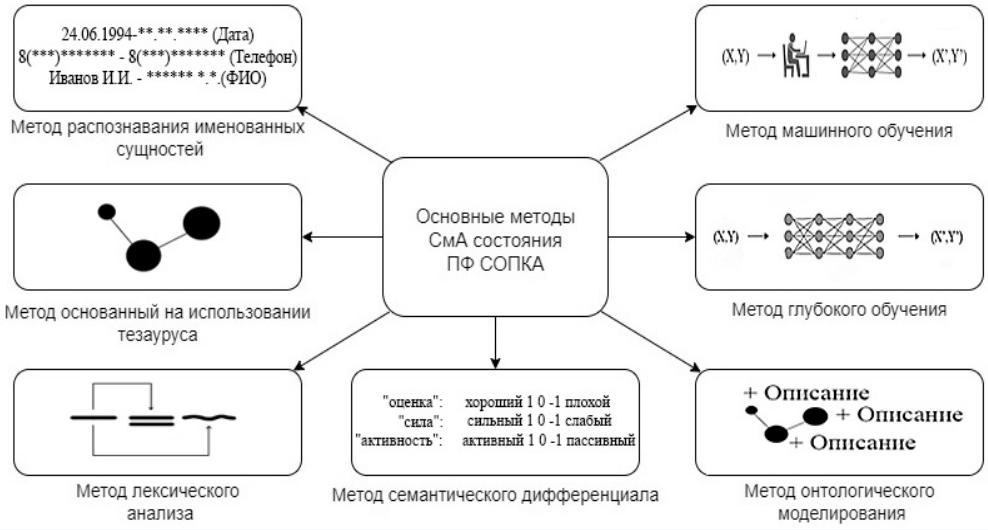

Рис. 1. Основные методы СмА состояния ПФ СОПКА
Fig. 1. Basic methods of SMA of the state of PF SOPKA

В целях определения наиболее эффективного метода, способного повысить оперативность исследуемого процесса, рассмотрим основные достоинства и недостатки существующих методов СмА состояния ПФ СОПКА (табл.1) [3-15]:

Таблица 1. Достоинства и недостатки существующих методов СмА состояния ПФ СОПКА
Table 1. Advantages and disadvantages of existing methods of SMA of the state of the PF SOPKA

Наименование метода СмА Name SMA	**Достоинства Advantages**	**Недостатки Disadvantages**
Распознавание Именованных сущностей Named Entity Recognition	позволяет идентифицировать различные именованные сущности посредством анализа ПД с учетом особенностей их написания; позволяет улучшить качество поисковых запросов, за счет выделяемых свойств ПД/ allows you to identify different named entities; allows you to improve the quality of search queries.	недостаточно эффективно классифицирует именованные сущности в тексте при обработке поступающих нестандартно расположенных ПД; недостаточный объем словаря синтаксических конструкций для различных языков/insufficiently classifies named entities in the text when processing incoming non-standardly located PD; insufficient volume of the dictionary for different languages
Тезаурус Thesaurus	учитывает семантические отношения между словами, что помогает уточнить семантические особенности и связи между ПД, улучшающие правильность понимания и интерпретацию текстовых данных/ takes into account the semantic relationships between words, which helps to clarify the semantic features and connections between PD, improving the correctness of understanding and interpretation of text data	представляет собой структуру устойчивых и заранее определенных семантических отношений между ПД, которые не учитывают контекстуальные изменения в значениях слов или различия при использовании в разных предметных областях; при изменении корпуса текстов и при повторной векторизации меняются числовые коэффициенты, что приводит к потребности дополнительных ресурсов/do not take into account contextual changes in the meanings of words or differences in use in different subject areas; when the text corpus changes and when re-vectorizing, the numerical coefficients change, which leads to the need for additional resources
Лексический анализ Lexical Analysis	позволяет разбить текст на отдельные слова, что упрощает дальнейший анализ и понимание семантики текста; по форме слова, по его сущности можно понять какое место, роль в предложении оно занимает/ allows you to break the text into separate words, you can understand what place, role in the sentence it occupies	допускает ошибки в распознавании некоторых символов, особенно при обработке текста с опечатками или при некорректности ПД; не информативен в аспекте выявления семантической информации в ПД, выделяя только их лексическое значение, роль в предложении/makes mistakes in recognizing some symbols, especially when processing text with typos or when the PD is incorrect; not informative in terms of identifying semantic information in the PD
Машинное обучение Machine Learning	высокая оперативность анализа большого объема ПД; высокая точность обработки ПД за счет обучения и дообучения модели СмА специалистом по ИБ/ high efficiency of analysis of a large volume of PD; high accuracy of PD processing due to training and additional training of the SMA model by an information security specialist	результаты анализа напрямую зависят от вида предобработанных ПД для обработки моделью машинного обучения; необходимость использования большого объема вычислительных, операционных и временных ресурсов/ the results of the analysis directly depend on the type of pre-processed PD for processing by the machine learning model; the need to use a large amount of computing, operational and time resources
Глубокое обучение Deep Learning	основано на нейронных сетях, которые могут определять сложные зависимости в ПД за счет подбора гиперпараметров и параметров машинного обучения; параметров машинного обучения, значения которых используются для управления процессом обучения нейронной сети/based on neural networks that	требует большого количества предобработанных ПД для обучения, что может быть затруднительно, особенно для некорректных ПД или небольших наборов ПД; необходимость использования большого объема вычислительных, операционных и временных ресурсов/requires a large number of pre-processed PDs for training, which can be difficult, especially for incorrect PDs or small sets of PDs; the need to use a large amount

Наименование метода СмА Name SMA	**Достоинства Advantages**	**Недостатки Disadvantages**
	can detect complex dependencies	of computing, operational and time resources
Онтологическое моделирование Ontology Modeling	гибкость настройки и повышение точности при формировании результатов поиска, позволяющее структурировать и описывать связи между понятиями в конкретной предметной области; онтология предоставляет организованную структуру и дает возможность учета семантических особенностей (холонимы, гипонимы и гиперонимы) для классификации ПД; содержит большой объем взаимосвязанных и упорядоченных ПД, что упрощает специалистом по ИБ их обработку; однозначность толкования относительно одного конкретного понятия между различными специалистами по ИБ/flexibility of customization and increased accuracy in generating search results; unambiguity of interpretation regarding one specific concept between different information security specialists	создание и поддержка онтологий представляет собой сложный и трудоемкий процесс; онтологии требуют аккуратной и точной формализации знаний группой специалистов по ИБ; интероперабельность, выраженная в сложности преобразования в единый формат получаемых ПД из различных источников знаний; отсутствие общего стандарта, универсальной методологии для создания онтологий/ creation and support of ontologies is a complex and labor-intensive process; ontologies require careful and precise formalization of knowledge by a group of information security specialists; interoperability, expressed in the complexity of converting the received PD from various knowledge sources into a single format; lack of a common standard, a universal methodology for creating ontologies
Семантический дифференциал Semantic Differential	относительная простота в использовании, заключающаяся в указании специалистами по ИБ своих предпочтений или оценок на шкалах, что делает данный метод относительно простым в применении и адаптации для различных задач анализа; шкалы метода семантического дифференциала могут охватывать различные аспекты семантики, что делает его относительно универсальным инструментом для анализа, подходящим для различных предметных областей/relative ease of use, the scales of the semantic differential method can cover various aspects of semantics, which makes it a relatively universal tool for analysis, suitable for various subject areas	анализ ПД ограничен выбранными специалистами по ИБ шкалами, что негативно отражается на последующей интерпретации и объективности полученных результатов анализа; интерпретация результатов анализа субъективна, и зависит от предположений специалиста по ИБ, особенно при анализе многофакторных ПД; сложность анализа, заключающаяся в необходимости применения как специальных статистических методов, так и дополнительного программного обеспечения при обработке ПД, что усложняет процесс анализа/ the analysis of PD is limited to scales selected by information security specialists, which negatively affects the interpretation and objectivity of the results; the interpretation of the analysis results is subjective and depends on the assumptions of the specialist; the complexity of the analysis, which consists in the need to use both special statistical methods and additional software when processing PD

Обсуждение результатов. Проведенный анализ существующих методов в рассматриваемой предметной области (табл.1) свидетельствует о том, что реализация СмА состояния ПФ СОПКА посредством какого-либо одного метода является недостаточно эффективным подходом.

При этом с учетом выявленных достоинств существующих методов СмА состояния ПФ СОПКА (табл. 1) повышение оперативности реализации исследуемого процесса может быть достигнуто за счет применения метода машинного обучения с использованием онтологического моделирования заданного объекта анализа. Кроме того, в целях возможной практической реализации СмА состояния ПФ СОПКА на основе метода машинного обучения с применением онтологического моделирования построим структурную модель соответствующей системы (рис. 2). Предназначение элементов структурной модели системы СмА состояния ПФ СОПКА (рис. 2) приведем в таб. 2.

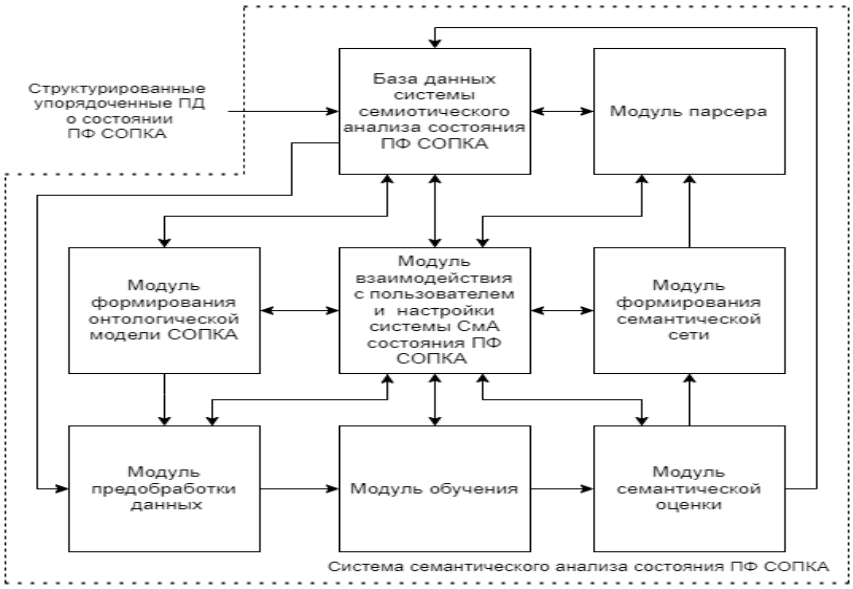

Рис. 2. Структурная модель предлагаемой системы СмА состояния ПФ СОПКА
Fig. 2. Structural model of the proposed SMA system for the state of PF SOPKA

Таблица 2. Предназначение элементов структурной модели системы СмА состояния ПФ СОПКА
Table 2. Purpose of the elements of the structural model of the system of the SMA state of the PF SOPKA

Наименование элемента **Name of the element**	**Функциональное предназначение элемента** **Functional purpose of the element**
База данных системы семиотического анализа состояния ПФ СОПКА Database of the semiotic analysis system of the state of the PF SOPKA	предназначена для длительного хранения и обработки: ПД о состоянии ПФ СОПКА, выделяемых из различных источников знаний и предоставляемых в виде соответствующих экзофреймовых структур; онтологической модели СОПКА; семантической сети, определяющей взаимосвязь между различными источниками знаний продукционно-фреймовых структур ПД о состоянии ПФ СОПКА и различных источников знаний СОПКА, онтологии СОПКА, семантической сети между ПД о состоянии ПФ СОПКА и их различными источниками знаний СОПКА/is intended for long-term storage and processing of: PD on the state of the SOPKA PF, extracted from various knowledge sources; ontological model of SOPKA; semantic network, SOPKA ontology, semantic network between PD on the state of the SOPKA PF
Модуль формирования онтологической модели СОПКА Module for forming the ontological model of SOPKA	предназначен для создания онтологической модели СОПКА в заданный момент времени ее функционирования/ is designed to create an ontological model of SOPKA at a given point in time of its operation
Модуль предобработки данных Data preprocessing module	предназначен для приведения структурированных упорядоченных ПД о состоянии ПФ СОПКА к виду, пригодному для обработки моделями машинного обучения/ is designed to bring structured ordered PD about the state of the PF SOPKA to a form suitable for processing by machine learning models
Модуль обучения Training module	предназначен для обучения и дообучения модели машинного обучения/ designed for training and retraining of machine learning models
Модуль семантической оценки Semantic evaluation module	предназначен для определения семантической близости между ПД из различных источников знаний о состоянии ПФ СОПКА/ is designed to determine the semantic proximity between PD from various sources of knowledge about the state of the PF SOPKA
Модуль парсера Parser module	предназначен для обработки линейной последовательности лексем (слов, токенов) естественного или формального языка, используемого специалистом по ИБ в заданной предметной области, с целью извлечения из базы данных системы семиотического анализа состояния ПФ СОПКА необходимой информации/is designed to process a linear sequence of lexemes (words, tokens) of a natural or formal language, in order to extract information from a database
Модуль взаимодействия пользователя и настройки системы СмА состояний ПФ СОПКА User interaction module and system configuration of the SMA states of the PF SOPKA	предназначен для настройки, управления системой и визуализации промежуточных результатов/ designed for setting up, managing the system and visualizing intermediate results
Модуль формирования семантической сети Module for the formation of a semantic network	предназначен для построения семантически связанной и упорядоченной структуры между ПД из различных источников знаний о состоянии ПФ СОПКА/ is designed to build a semantically linked and ordered structure between PD from various sources of knowledge about the state of the PF SOPKA

Вывод. Таким образом, проведенный анализ в заданной предметной области свидетельствует о том, что наиболее перспективным решением к семантическому анализу состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак является использование метода машинного обучения с применением онтологического моделирования.

Созданная в ходе исследования и представленная в статье модель системы семантического анализа способна в полной мере обеспечить анализ состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак.

Исходя из вышеуказанного, проведенный анализ существующих методов семантического анализа свидетельствует о необходимости дальнейшего исследования в разработке предлагаемой системы семантического анализа состояния процесса функционирования системы обнаружения, предупреждения и ликвидации последствий компьютерных атак.

Библиографический список:

1. Коноваленко С.А., Королев И.Д., Шабля В.О. Анализ процесса функционирования ведомственного сегмента системы обнаружения, предупреждения и ликвидации последствий компьютерных атак на критическую информационную инфраструктуру Вооруженных Сил Российской Федерации // Информационная безопасность – актуальная проблема современности. Совершенствование образовательных технологий подготовки специалистов в области информационной безопасности: материалы XXIII Всерос. межведомст. НТК, г. Краснодар 2021 г. / отв. ред. д.т.н., проф. А.В. Крупенин. – Краснодар: КВВУ, 2021. – Т. 2. – С. 80-90.
2. Шабля В.О., Коноваленко С.А., Едунов Р.В. Анализ процесса функционирования SIEM-систем // E-scio [Электронный ресурс]: Электронное периодическое издание "E-scio.ru" - Эл № ФС77-66730 - Режим доступа: http:e-scio.ru/wp-content/uploads/2022/05/Шабля В.О.,КоноваленкоС. А.,Едунов Р..-В.pdf
3. Коноваленко С.А., Шабля В.О., Титов Г.О. Анализ методов контроля состояния процесса функционирования сложных технических систем [Электронный ресурс]//Наукосфера.-2021,- № 12(2), С. 234.
4. Кульневич А.Д., Кошечкин А.А., Карев С.В., Замятин А.В. Подход к распознаванию именованных сущностей на примере технологических терминов в условиях ограниченной обучающей выборки. Вестник Томского государственного университета. Управление, вычислительная техника и информатика, (58), с. 71-81.
5. Лесников, С.В. Тезаурус как отражение системности языка. Вестник Челябинского государственного университета, (28), С. 52-61.
6. Осокина С.А. Сетевая модель языкового тезауруса: особенности построения. Сибирский филологический журнал, (3), С. 191-198.
7. Лазутченкова, Е.А. Прагматический анализ в лексической семантике. Полилингвиальность и транскультурные практики, (1), С. 62-65.
8. Боброва М.Б., Мастилин А.Е. Машинное обучение в кибербезопасности. Научные междисциплинарные исследования, (2), С. 24-29.
9. Магжанова А.Т. Интеграция информационных источников с использованием кластер- анализа по схеме машинного обучения без учителя. Теория и практика современной науки, (6 (24)), С. 1037-1040.
10. Максютин П.А., Шульженко С.Н. Обзор методов классификации текстов с помощью машинного обучения. Инженерный вестник Дона, (12 (96)), С. 1-9.
11. Жиленков А.А., Силкин А.А., Серебряков М.Ю., Колесова С.В. Сравнительный анализ систем глубокого обучения с подкреплением и систем обучения с учителем. Известия Тульского государственного университета. Технические науки, (10), С. 109-112.
12. Столяров А.С., Раджабов Т.Р. Развитие ИИ, глубокое и машинное обучение. Теория и практика современной науки, (8 (38)), С. 70-80.
13. Папуша С.И. Онтология и графовые базы данных. Проблемы экономики и юридической практики, (3), С. 268-272.
14. Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д. Онтологии и тезаурусы: модели, инструменты, приложения: учебное пособие [Электронный ресурс] / Б.В. Добров, В.В. Иванов, Н.В. Лукашевич, В.Д. Соловьев. — 2-е изд. (эл.) — Электрон, дан. и прогр. (3 Мб.) — М: Интернет-Университет Информационных Технологий; Саратов: Вузовское образование, 2017.
15. Хекало Т.В. Изучение личностных смыслов физико-химических объектов методом семантического дифференциала. Вопросы психолингвистики, 2016, С. 256-265.

References:

1. Konovalenko S.A., Korolev I.D., Shablya V.O. Analysis of the functioning of the departmental segment of the system for detecting, preventing and eliminating the consequences of computer attacks on the critical information infrastructure of the Armed Forces of the Russian Federation. Information security – an urgent problem of our time. Improving educational technologies for training specialists in the field of information security: materials of the XXIII Allround. interagency. NTC, Krasnodar, 2021 / ed., Doctor of Technical Sciences, prof. A.V. Krupenin. – Krasnodar: KVVU, 2021; 2: 80-90. (In Russ)
2. Shablya V.O., Konovalenko S.A., Edunov R.V. Analysis of the process of functioning of SIEM systems // Esco [Electronic resource]: Electronic periodical "E-scio.ru " - E-mail no. FS77-66730 - Access mode: http:e-scio.ru/wp-content/uploads/2022/05/Шабля-В.-О.-Коноваленко-С.-А.-Едунов-Р.-В.pdf (In Russ)
3. Konovalenko S.A., Shablya V.O., Titov G.O. Analysis of methods for monitoring the state of the process of functioning of complex technical systems [Electronic resource]. *The science sphere*. 2021;12 (2): 234. (In Russ)
4. Kulnevich A.D., Koshechkin A.A., Karev S.V., Zamyatin A.V. An approach to the recognition of named entities by the example of technological terms in a limited training sample. Bulletin of Tomsk State University. Management, Computer Engineering and Computer Science, (58):71-81. (In Russ)
5. Lesnikov, S.V. Thesaurus as a reflection of the consistency of language. Bulletin of the Chelyabinsk State University, 28; 52-61. (In Russ)
6. Osokina S.A. The network model of the language thesaurus: features of construction. Siberian Journal of Philology, 3; 191-198. (In Russ)
7. Lazutchenkova, E. A. Pragmatic analysis in lexical semantics. Polylingualism and Transcultural Practices, 1: 62-65. (In Russ)
8. Bobrova M.B., Mastilin A.E. Machine learning in cybersecurity. Scientific Interdisciplinary Research, 2: 24-29. (In Russ)
9. Magzhanova A.T. Integration of information sources using cluster analysis according to the machine learning scheme without a teacher. Theory and Practice of Modern Science, 6 (24:1037-1040. (In Russ)
10. Maksyutin P.A., Shulzhenko S.N. Review of text classification methods using machine learning. Engineering Bulletin of the Don, 12 (96):1-9. (In Russ)
11. Zhilenkov A.A., Silkin A.A., Serebryakov M.Yu., Kolesova S.V. Comparative analysis of systems deep reinforcement learning and teacher-led learning systems. Proceedings of Tula State University. Technical Sciences, 10:109-112. (In Russ)
12. Stolyarov A.S., Rajabov T.R. The development of AI, deep and machine learning. Theory and Practice of Modern Science, 8 (38): 70-80. (In Russ)
13. Papusha S.I. Ontology and graph databases. Problems of Economics and legal practice, (3), pp. 268-272. (In Russ)
14. Dobrov B.V., Ivanov V.V., Lukashevich N.V., Solovyov V. D. Ontologies and thesauruses: models, tools, applications: textbook [Electronic resource] / B. V. Dobrov, V.V. Ivanov, N.V. Lukashevich, V.D. Solovyov. — 2nd ed. (electronic) — Electron, dan. and progr. (3 Mb.) — Moscow: Internet University of Information Technologies; Saratov: University Education, 2017. (In Russ)
15. Hekalo T.V. The study of personal meanings of physico-chemical objects by the method of semantic differential. Questions of Psycholinguistics, 2016:256-265. (In Russ)

Сведения об авторах:
Шабля Владимир Олегович, аспирант; ne.404@mail.ru
Коноваленко Сергей Алесандрович, докторант, кандидат технических наук, доцент; konovalenko_rcf@mail.ru
Орлов Егор Олегович, студент 4 курса; egor535553@mail.ru
Information about authors:
Vladimir O. Shablya, Postgraduate Student; ne.404@mail.ru
Sergey A. Konovalenko, Doctoral Student, Cand. Sci. (Eng), Assoc. Prof.; konovalenko_rcf@mail.ru
Egor O. Orlov, 4th year Student; egor535553@mail.ru
Конфликт интересов/Conflict of interest.
Авторы заявляют об отсутствии конфликта интересов/The authors declare no conflict of interest.
Поступила в редакцию/Received 21.05.2024.
Одобрена после рецензирования/ Reviced 20.06.2024.
Принята в печать/Accepted for publication 09.10.2024
.