ISSN 2686-679X

ВЕСТНИК РГГУ

Серия
«Информатика.
Информационная безопасность.
Математика»

Научный журнал

RSUH/RGGU BULLETIN

"Information Science.
Information Security. Mathematics"
Series

Academic Journal

Основан в 2018 г.
Founded in 2018

2
2024

VESTNIK RGGU. Seriya “Informatica. Informacionnaya bezopasnost. Matematica”

RSUH/RGGU BULLETIN. “Information Science. Information Security. Mathematics” Series
Academic Journal

There are 4 issues of the printed version of the journal a year.

Founder and Publisher
Russian State University for the Humanities (RSUH)

RSUH/RGGU BULLETIN. “Information Science. Information Security. Mathematics” series is included: in the Russian Science Citation Index; in the List of leading scientific magazines journals and other editions for publishing PhD research findings peer-reviewed publications fall within the following research area:

1.1.6. Computational Mathematics (physical and mathematical sciences)

2.3.6. Information security methods and systems, information security (technical science)

2.3.8. Informatics and information processes (technical science)

Objectives and areas of research

RSUH/RGGU BULLETIN. “Information Science. Information Security. Mathematics” series publishes the results of research by scientists from RSUH and other universities and other Russian and foreign academic institutions. The areas covered by contributions include theoretical and applied computer science, up-to-date IT, means and technologies of information protection and information security as well as the issues of theoretical and applied mathematics including analytical and imitation models of different processes and objects. Special emphasis is put on articles and reviews covering research in indicated directions in the areas of social and humanitarian problems and also issues of personnel training for these directions.

RSUH/RGGU BULLETIN. “Information Science. Information Security. Mathematics” series is registered by Federal Service for Supervision of Communications Information Technology and Mass Media. 25.05.2018, reg. No. FS77-72977

Editorial staff office: 6, Miusskaya sq., Moscow, Russia, 125047
e-mail: grnat@rambler.ru

ВЕСТНИК РГГУ. Серия «Информатика. Информационная безопасность. Математика»
Научный журнал

Выходит 4 номера печатной версии журнала в год.

Учредитель и издатель – Российский государственный гуманитарный университет (РГГУ)

ВЕСТНИК РГГУ, серия «Информатика. Информационная безопасность. Математика», включен: в систему Российского индекса научного цитирования (РИНЦ); в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук по следующим научным специальностям и соответствующим им отраслям науки:

1.1.6. Вычислительная математика (физико-математические науки)

2.3.6. Методы и системы защиты информации, информационная безопасность (технические науки)

2.3.8. Информатика и информационные процессы (технические науки)

Цели и область

В журнале «Вестник РГГУ», серия «Информатика. Информационная безопасность. Математика», публикуются результаты научных исследований ученых и специалистов РГГУ, а также других университетов и научных учреждений России и зарубежных стран. Направления публикаций включают теоретическую и прикладную информатику, современные информационные технологии, методы, средства и технологии защиты информации и обеспечения информационной безопасности, а также проблемы теоретической и прикладной математики, включая разработку аналитических и имитационных моделей процессов и объектов различной природы. Особое внимание уделяется статьям и обзорам, посвященным исследованиям по указанным направлениям в области социальных и гуманитарных проблем, а также вопросам подготовки кадров по соответствующим специальностям для данных направлений.

ВЕСТНИК РГГУ. Серия «Информатика. Информационная безопасность. Математика», зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 25.05.2018 г., регистрационный номер ПИ № ФС77-72977.

Адрес редакции: 125047, Россия, Москва, Миусская пл., 6

Электронный адрес: grnat@rambler.ru

Учредитель и издатель
Российский государственный гуманитарный университет (РГГУ)

Ответственный за выпуск:

Н.В. Гришина, кандидат технических наук, доцент,
Российский государственный гуманитарный университет (РГГУ)

CONTENTS

Information Science

Igor V. Timoshenko
The normative base of information retrieval thesauri. History and current development trends 8

Andrei P. Titov
Software implementation of the Co-Active Neuro-Fuzzy Inference System 26

Elena O. Shershneva, Yurii Yu. Karyuchin
Analysis and optimization of the reverse product data exchange system using the "ST Chicago" functional software package 44

Information Security

Leonid E. Alekseev, Aleksei E. Samokhvalov, Evgenia R. Smolina
The system of intellectual analysis of texts and identification of elements of information confrontation 58

Natalia V. Grishina
Analysis of approaches to investigating information security incidents 73

Nikolai A. Ignat'ev, Erkin R. Navruzov
On patterns in detecting denial of service attacks in computer networks 83

Andrei S. Kuznetsov, Andrei E. Krasnov
Information support for pulse control of the stability of information security systems 99

СОДЕРЖАНИЕ

Информатика

Игорь В. Тимошенко
Нормативная база информационно-поисковых тезаурусов: история и современные тенденции развития 8

Андрей П. Титов
Программная реализация модели Co-Active Neuro-Fuzzy Inference System 26

Елена О. Шершнева, Юрий Ю. Карючин
Анализ и оптимизация существующей системы обмена данными реализации продукции с применением функционала программного комплекса “ST Chicago” 44

Информационная безопасность

Леонид Е. Алексеев, Алексей Э. Самохвалов, Евгения Р. Смолина
Система интеллектуального анализа текстов и выявления элементов информационного противоборства 58

Наталия В. Гришина
Анализ подходов к расследованию инцидентов информационной безопасности 73

Николай А. Игнатьев, Эркин Р. Наврузов
О закономерностях при обнаружении атак «отказ в обслуживании» в компьютерных сетях 83

Андрей С. Кузнецов, Андрей Е. Краснов
Информационное обеспечение импульсного управления устойчивостью систем информационной безопасности 99

УДК 027.7:004.5
DOI: 10.28995/2686-679X-2024-2-8-25

Нормативная база информационно-поисковых тезаурусов: история и современные тенденции развития

Игорь В. Тимошенко
Московский государственный институт культуры, Москва, Россия;
Государственная публичная научно-техническая библиотека России, Москва, Россия, timigor@gpntb.ru

Аннотация. В статье представлен обзор основных концепций построения информационно-поисковых тезаурусов, применяемых для индексирования информационных ресурсов библиотек. Механизмы информационного поиска развивались исторически, используя метаданные – поисковые индексы для определения тематического соответствия проиндексированных информационных ресурсов поисковым запросам, исходя из их описания. Большое распространение в последние десятилетия получили словарные методы информационного поиска. В статье обсуждается история развития информационно-поисковых тезаурусов – специализированных словарей, применяемых для индексирования ресурсов в информационно-поисковых системах начиная с 1960-х гг., а также их роль в повышении точности и эффективности поиска информации. Рассмотрено влияние развития электронных средств коммуникации и сети Интернет на принципы формирования и структуру информационно-поисковых тезаурусов. Показано развитие нормативной базы тезаурусов, применяемых в библиотечной области, и ее связь со стандартами Интернета в концепции Семантической паутины. Основные этапы развития нормативной базы в нашей стране показаны как на международном, так и на национальном уровне. Отмечается важность изучения опыта других организаций в области стандартизации для дальнейшего развития нормативной базы индексирования библиотечных ресурсов.

Ключевые слова: тезаурус, индексирование, информационные ресурсы, информационный поиск, Интернет, Семантическая паутина

Для цитирования: *Тимошенко И.В.* Нормативная база информационно-поисковых тезаурусов: история и современные тенденции развития // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 8–25. DOI: 10.28995/2686-679X-2024-2-8-25

The normative base of information retrieval thesauri. History and current development trends

Igor V. Timoshenko
Moscow State Institute of Culture, Moscow, Russia;
Russian National Public Library for Science and Technology, Moscow, Russia, timigor@gpntb.ru

Abstract. The article presents a review of the main concepts of building information retrieval thesauri used for indexing the information resources of libraries. The author shows historical development of the information retrieval mechanisms, using metadata as search term to determine the thematic correspondence of indexed information resources to search queries, based on their description. Dictionary-based methods of information retrieval became widely used in recent decades. The author discusses history of the information retrieval thesauri development specialized dictionaries used to indexing resources in information retrieval systems, starting in the 1960s, for improving the accuracy and efficiency of information search. The impact of the development of electronic means of communication and the Internet on the principles of formation and structure of information retrieval thesauri is considered.

The article shows development in the thesauri normative base used in the libraries and its connection with the Internet standards in the concept of the Semantic Web. Main stages in development of the regulatory framework are revealed, both at the international level and at the national, in our country. The author notes importance of studying the experience of other organizations in the field of standardization for the further development of the regulatory normative base for indexing library resources.

Keywords: thesaurus, indexing, information resources, information retrieval, Internet, Semantic web

For citation: Timoshenko, I.V. (2024), "The normative base of information retrieval thesauri. History and current development trends", *RSUH/RGGU Bulletin. "Information science. Information security. Mathematics" Series,* no. 2, pp. 8–25, DOI: 10.28995/2686-679X-2024-2-8-25

Введение

Электронные средства коммуникации и глобальная сеть Интернет значительно трансформировали структуру библиотечных фондов и методы библиографирования информационных ресурсов библиотек. Появилась тенденция включения в состав библиотечных фондов нетекстовых документов, таких как аудио, изображения, видео и мультимедиа. Библиотеки работают с электронными каталогами, которые представляют собой организованные системы библиографических записей, содержащие описания всех информационных ресурсов фонда, независимо от их формы. Сам электронный каталог является основным информационным ресурсом библиотеки, представляющим собой модель библиотечного фонда, обеспечивающую идентификацию документов фонда во всех библиотечных процессах через библиографические данные, представленные в его записях. Основной функцией электронного каталога библиотеки является обеспечение поиска информации по запросам пользователей. Механизмы поиска информации развивались на протяжении всей истории библиотек, при этом общий подход всегда заключался в индексировании информационных ресурсов, т. е. в присвоении ресурсам системы метаданных – поисковых индексов, облегчающих определение их соответствия поисковым запросам. В данной статье рассмотрены основные концепции поискового индексирования и формирования информационно-поисковых тезаурусов, применяемых в библиотеках, в связи с развитием механизмов информационного поиска Интернета в концепции Семантической паутины.

Основные подходы к индексированию информационных ресурсов

Исторически сложились два подхода к формированию структуры поисковых индексов для индексирования информационных ресурсов: классификационный и полнотекстовый.

Классификационный подход основан на предварительной разработке схемы знаний, как правило, имеющей иерархическую структуру и относящейся к конкретной предметной области. Все элементы такой схемы кодируются, коды используются для формирования поисковых индексов, которые присваиваются информационным ресурсам. Каждый код уникально соответствует разделу знаний в рамках применяемой схемы и формируется исходя из анализа и оценки информационного ресурса. В дальнейшем пользовательский

поиск осуществляется по структуре схемы знаний, без привлечения проиндексированных ресурсов.

Полнотекстовый подход заключается в оценке тематики ресурса на основании анализа структуры и элементов самого ресурса. Такой подход отличает принципиальная возможность достижения абсолютной точности в определении соответствия ресурса информационной потребности пользователя, но также и значительно большая трудоемкость, что не позволяет его широко применять при ручном поиске в сколько-нибудь обширных коллекциях информационных ресурсов. Снижение трудоемкости поиска возможно путем предварительной обработки текста, которая заключается в выделении его основных смысловых элементов и изложении их в сокращенном виде. На практике применяются три формы такой обработки: реферирование, аннотирование и выделение из текста ключевых слов. В дальнейшем пользовательский поиск осуществляется путем анализа сформированных элементов, также без привлечения исходных информационных ресурсов.

Классификационные методы поиска получили широкое распространение в докомпьютерную эпоху. В конце XIX – первой половине XX в. был разработан ряд универсальных классификационных систем, применявшихся в поисковых системах библиотек, основанных на карточных каталогах. Информационный поиск осуществлялся по кодам классификационных систем, точность поиска зависела от правильности присвоения кодированных индексов информационным ресурсам, что при условии тематической многоаспектности ресурсов представляет собой нетривиальную задачу и требует определенной квалификации, навыков кодирования. Полнотекстовый метод поиска в карточных каталогах носил вспомогательный, уточняющий характер и выражался в анализе аннотаций ресурсов, приводимых в каталожных карточках.

Развитие словарных методов индексирования

Вторая половина XX в. характеризуется бурным ростом количества информации, циркулирующей в обществе, отражающей как культурные, так и научно-технические аспекты деятельности людей. Внедрение в информационные сферы деятельности электронно-вычислительной техники как инструмента, многократно повышающего производительность человека в процессах обработки информации, прежде всего в задачах информационного поиска, произвело в ней революционные изменения. Прежде всего применение методов машинной обработки информации суще-

ственно изменило преимущественный подход к методам информационного поиска. Развитие методов машинной обработки информации привело к созданию баз данных, появлению электронных документов, электронных средств распространения информации. Моделирование процессов организации и поиска больших массивов информации в электронной среде привело к появлению теории баз данных, основанной на реляционной модели, а также теории информационного поиска, развивающей модели поиска в словарных поисковых системах с применением информационных запросов, выраженных на естественных языках.

Преимущественное развитие словарных методов поиска, основанных на информационно-поисковых языках (ИПЯ) ключевых слов, связано с развитием методов машинной обработки текстов. Процесс формирования и присвоения кодов универсальных классификационных систем в большей степени носит эвристический характер, и до сегодняшнего дня задача его автоматизации в полном объеме не решена. По этой причине формирование и использование классификационных кодов увеличивает затраты ручного труда на предварительную обработку (предкоординатное индексирование) и последующий поиск информационных ресурсов, поэтому не пользуется популярностью ни у библиотечного персонала, ни у пользователей, которые предпочитают выражать свои запросы на естественном языке.

Задача сопоставления ключевых слов в информационном запросе, выраженном на естественном языке с ключевыми словами, которыми проиндексированы ресурсы информационно-поисковых систем с применением машинной обработки, удовлетворительно решается четкими алгоритмическими методами, реализуемыми на компьютерах сравнительно небольшой вычислительной мощности, и ее реализация была доступна уже для ЭВМ поколений 1960–1970-х гг.

Вместе с тем простая концепция поиска по ключевым словам сталкивается с проблемой вариативности естественных языков – одно и то же понятие может быть выражено множеством различных слов и словосочетаний. Идеальный результат поиска по ключевым словам достижим при условии применения каждого конкретного термина только к одному понятию, для этого нужно приучить людей всегда использовать одни и те же термины для одних и тех же понятий. Ввиду практической нереализуемости этой идеи в задачах поиска информации широкое распространение получил тезаурусный подход. Применение тезауруса кратно уменьшает количество терминов, применяемых в процессах информационного поиска. Тезаурус помогает выбрать однозначный термин для выражения

понятия, как для индексирования искомого ресурса, так и для выражения поискового запроса.

Создание информационно-поискового тезауруса (ИПТ) требует предварительной работы, связанной с определением терминосистемы конкретной предметной области с последующим подбором и группировкой вокруг установленных терминов их синонимов, извлекаемых из корпуса текстов, относящихся к этой области. Общая структура тезауруса, с учетом возможных иерархических связей между определяемыми понятиями, может быть выражена моделью, аналогичной модели иерархических классификационных рубрикаторов или основных таблиц универсальных классификаций. Это обстоятельство, в большой мере, стирает принципиальные различия между двумя подходами применения поисковых индексов, традиционно противопоставляемых друг другу, предкоординатным, классификационным индексированием и координатным – ключевыми словами, с использованием ИПТ.

Развитие нормативной базы информационно-поисковых тезаурусов

Широкое внедрение ЭВМ в информационные сферы деятельности и развитие словарных методов информационного поиска в мире привело к появлению уже в 1960-х гг. большого числа тезаурусов, применяемых для индексирования ресурсов в крупных отраслевых информационно-поисковых системах (ИПТ). Первым наиболее известным из них был «Тезаурус инженерных и научных терминов» (TEST), впервые опубликованный в 1964 г. и переизданный в 1967 г. [Thesaurus 1967]. Тезаурус TEST был разработан Объединенным советом инженеров при участии Министерства обороны США, он содержал более 17 000 дескрипторов, представленных в алфавитном порядке, которые объединяли более 150 000 терминов, с разъяснениями, как они могут быть использованы для поиска информации. Важной особенностью TEST было то, что он включал раздел с установленными правилами и соглашениями о своей структуре, которые послужили основой для дальнейших работ по стандартизации принципов построения информационно-поисковых тезаурусов. Аналогичный русскоязычный «Тезаурус научно-технических терминов» [Шемакин 1972] был опубликован в СССР в 1972 г.

Альтернативный подход к построению тезауруса был реализован в «Тезаурусе по искусству и архитектуре» (AAT), впервые опубликованном в 1970 г., который представляет собой систему

фасетной классификации. Сегодня словарь включает более 55 000 дескрипторов, объединяющих более 130 000 лексических единиц. Лексические единицы тезауруса распределены по семи иерархическим аспектам [Petersen 1994].

На основании опыта, полученного при разработке крупнейших тезаурусов, усилиями рабочей группы экспертов технического комитета ИСО/ТК 46 в 1971 г. был разработан первый международный документ стандартизации "ISO/TC 46/T Recommendation 1041E of August 3, 1971", устанавливающий принципы создания и применения тезаурусов на международном уровне. Этот документ послужил основой для ряда национальных документов стандартизации, в том числе в нашей стране.

В СССР к концу 1960-х гг. существовало уже порядка ста крупных автоматизированных информационно-поисковых систем (ИПС), использующих различные тезаурусы как ИПЯ координатного индексирования [Данилов, Уманский 1969]. Для решения проблемы их совместимости в 1965 г. Государственный комитет по науке и технологиям СССР (ГКНТ СССР) инициировал проект создания «Единой системы классификации печатных изданий и документальных материалов» (ЕСКПИДМ). Ключевым итогом работы над этим проектом, реализованным в 1965–1970 гг. во ВНИИКИ Госстандарта СССР, стал вывод о невозможности создания единой обязательной системы классификации. Решение проблемы языковой совместимости возможно путем систематизации существующих ИПЯ с созданием на их базе Комплекса средств индексирования научно-технической информации (КСИНТИ) [Антопольский, Вайсберг, Зарувинская 1976]. Цель создания комплекса заключалась в разработке и практической проверке методов унификации и максимально автоматизированного анализа основных тезаурусов, используемых в Государственной автоматизированной системе научно-технической информации (ГАСНТИ), для их согласованного ведения и гармонизированного сопоставления. На основе опыта, полученного в ходе реализации проекта, к 1973 г. Государственным комитетом стандартов Совета Министров СССР (Госстандарт СССР), а также с учетом рекомендаций, установленных в документе "SO/TC 46/T Recommendation 1041E of August 3, 1971", был разработан и утвержден ГОСТ 18383-73 «Тезаурус информационно-поисковый. Общие положения. Форма представления»[1]. Стандарт устанавливал самые общие требования к построению ИПТ: обя-

[1] ГОСТ 18383-73 Тезаурус информационно-поисковый. Общие положения. Форма представления. URL: https://docs.cntd.ru/document/464617921?ysclid=lroymb74i8432294033 (дата обращения 12.01.2024).

зательные элементы структуры, порядок расположения терминов дескрипторной статьи и уточнения в случае их омонимии или полисемии. Стандарт распространялся на тезаурусы для обработки и поиска научно-технической информации в отраслевых ИПС и был обязателен для применения на территории СССР. Также стандарт определял необходимость регистрации в нашей стране тезаурусов, в соответствии с порядком, установленным ГКНТ СССР.

В 1976 г. в рамках реализации проекта Автоматизированной системы ведения информационных языков (АСВИЯ) задача экспертизы, регистрации информационно-поисковых языков была закреплена за ВИНИТИ. Совместно с ВНИИКИ был разработан и согласован с соисполнителями нормативный документ «Порядок экспертизы и регистрации информационно-поисковых тезаурусов» [Тимошенко 2023, с. 97–98].

На основании ранее опубликованного документа “T Recommendation 1041E”, международная организация по стандартизации ИСО к 1974 г. разработала и опубликовала международный стандарт ИСО 2788:1974 «Документация. Руководство по созданию и развитию одноязычных тезаурусов», в котором были более детально представлены принципы и правила, установленные ранее в рекомендациях и в тезаурусе TEST. В 1986 г. стандарт был переиздан, а также в этот период появился стандарт ИСО 5964:1985 «Документация. Руководство по созданию и развитию многоязычных тезаурусов[2], который расширял область применения основных принципов, установленных в ИСО 2788 на многоязычные тезаурусы.

Главная цель ИСО 2788 декларируется как систематизация понятий и их взаимосвязей. Для указания иерархических отношений между терминами в стандарте рекомендуется использовать теги, изначально предложенные в правилах тезауруса TEST: BT (более широкий термин), NT (более узкий термин), RT (родственный термин) и т. д. Принципиальным моментом, устанавливаемым правилами первых стандартов ИСО, являлось отождествление «понятия» и «термина», его выражающего. В последней редакции стандарта ИСО 2788:1986 эта особенность явно указывается во введении: «Для практических целей “термин” и “понятие” иногда используются взаимозаменяемо»[3].

---

[2] ISO 5964:1985 Documentation Guidelines for the establishment and development of multilingual thesauri. URL: https://www.iso.org/standard/12159.html (дата обращения 12.01.2024).

[3] ISO 2788:1986 Documentation Guidelines for the establishment and development of monolingual thesauri, URL: https://www.iso.org/standard/7776.html (дата обращения 12.01.2024).

Такое отождествление является характерной чертой и для родственного российского стандарта ГОСТ 7.25-2001 «Тезаурус информационно-поисковый одноязычный»[4], впервые опубликованного в 1980 г. и заменившего ГОСТ 18383-73. В стандарте установлен более широкий набор русскоязычных и символьных обозначений связей между лексическими единицами тезауруса, включающий иерархические и ассоциативные связи, аналогичные англоязычным обозначениям. Стандарт определяет возможность указания иных парадигматических связей между лексическими единицами тезауруса, что дает пользователю информацию о существовании нескольких понятий, выражаемых одним термином, но выбор термина остается за пользователем. Правила построения многоязычных тезаурусов установлены в ГОСТ 7.24-2007 «Тезаурус информационно-поисковый многоязычный», впервые опубликованном в 1980 г. Стандарт расширял действие правил, установленных в ГОСТ 7.25 на многоязычные тезаурусы, а также он был разработан с учетом основных нормативных положений международного стандарта ИСО 5964:1985.

Эволюция концепции построения тезаурусов

Основное назначение тезауруса – помочь пользователю выбрать правильный термин, отражающий заданное понятие, при этом следует заметить, что понятие исходно существует только в представлении пользователя и, в общем случае, не зависит от термина и языка, на котором он выражен. Необходимость представления понятия в форме кода или термина возникает в акте коммуникации, который может быть реализован не только человеком, но и автоматизированными системами.

Совмещение представления о «термине» и «понятии» не имеет большого значения в случае, когда подбор термина из тезауруса осуществляется при участии человека. Различие между термином и понятием человек определяет интуитивно, исходя из контекста коммуникации, и, как правило, может адекватно интерпретировать результат поиска. В случае машинной обработки информационная система должна иметь более четкие условия для интерпретации используемых терминов. Существующая проблема

[4] ГОСТ 7.25–2001 Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. URL: https://docs.cntd.ru/document/1200025969 (дата обращения 12.01.2024).

вариативности естественных языков может приводить к ошибкам и ложным выводам при выборе терминов, при формальной их правильности, что можно проиллюстрировать на примере из силлогистики Аристотеля:

«Аристотель – грек;

грек – слово из четырех букв;

Аристотель – слово из четырех букв».

Опережающее развитие автоматизированных информационных технологий в последние десятилетия создало новые условия и выдвинуло новые требования к лингвистическим средствам коммуникаций. Компьютеризация информационных видов деятельности, развитие электронных средств коммуникации и глобальной сети Интернет привели к появлению концепций Семантической паутины и Интернета вещей, а также феномена искусственного интеллекта и интеллектуальных информационных систем. Автоматизация процессов семантической обработки информационных ресурсов, в рамках которых подразумевается возможность извлечения информации из множества сетевых ресурсов не только с участием человека, диктует необходимость более четкого определения «понятия» как объекта и формализацией выбора адекватного ему термина для индексирования ресурса, что возможно путем развития моделей представления данных, устанавливающих понятия не только в широком контексте определяющих его терминов тезауруса, но и в контексте разнородных словарей лексических единиц, применяемых в задачах информационного поиска.

Развитие информационных сервисов Интернета способствовало появлению большого разнообразия контролируемых словарей и видов лексики, используемых для навигации, фильтрации, поиска информации в разнообразных сетевых ресурсах. Кроме потребностей семантической обработки самих информационных ресурсов, появилась потребность в определении требований к функциональной совместимости между собой различных тезаурусов, а также их совместимости с другими словарями, применяемыми в различных информационных системах Интернета. Только реализация такой совместимости может обеспечить возможность семантической обработки всей совокупности информационных ресурсов, представленных в Интернете, в задачах информационного поиска. Реализация этой идеи возможна путем эволюции Интернета в направлении концепции Семантической паутины[5].

[5] W3C Semantic Web Frequently Asked Questions. URL: https://www.w3.org/RDF/FAQ/ (дата обращения 12.01.2024).

Семантическая паутина концептуально представляет собой универсально общедоступную платформу, позволяющую обмен информационными ресурсами и их обработку как людьми, так и техническими средствами информационных систем. Развитие в этом направлении ведется по инициативе международной организации W3C (World Wide Web Consortium). Концепция является надстройкой над существующей Всемирной паутиной (WWW) и базируется на технологии уникальной идентификации URI (Uniform Resource Identifier)[6] и модели представления данных RDF (Resource Description Framework)[7], позволяющей описывать структуру семантической сети в виде графа, образующего онтологию, состоящую из индексированных информационных ресурсов. Онтологии позволяют объединять информационные ресурсы в концептуальные модели, исходя из информационных потребностей конкретной предметной области. Примером таких моделей могут служить концептуальные модели библиографических данных[8] IFLA. Индексирование информационных ресурсов как компонентов онтологий знаний позволяет формировать базы знаний для получения логических выводов по правилам математических логик в задачах информационного поиска, реализуемых в интеллектуальных информационных системах.

Современное состояние нормативной базы информационно-поисковых тезаурусов

Необходимым условием реализации концепции Семантической паутины является переход от существующей технологии индексирования информационных ресурсов, ориентированных на традиционные «бумажные» технологии, на новую, отвечающую требованиям электронной коммуникационной среды и реализующую ее возможности. Разработка нормативной базы новой технологии велась с конца 1990-х гг. Началом работ в этом направлении

---

[6] Uniform Resource Identifier (URI) // Generic Syntax. URL: https://www.ietf.org/rfc/rfc3986.txt (дата обращения 12.01.2024)

[7] Resource Description Framework (RDF): Concepts and Abstract Syntax // W3C Recommendation 25 February 2014. URL: https://www.w3.org/TR/rdf-concepts/ (дата обращения 12.01.2024).

[8] IFLA Library Reference Model: A Conceptual Model for Bibliographic Information//IFLA.URL:https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf (дата обращения 12.01.2024).

можно считать проекты DESIRE II, LIMBER и SWAD-Europe[9], в рамках которых разрабатывались методы представления тезаурусов на языке RDF. С 2004 г. работа была продолжена рабочей группой консорциума W3C. Результатом усилий стала разработка стандарта W3C: “Simple Knowledge Organization System” (SKOS), официальный релиз которого был опубликован в 2009 г.[10]

Стандарт SKOS изначально рассматривался как связующее звено между традиционными системами организации знаний, применяемыми в библиотечно-информационной области, такими как тезаурусы, и разнообразными веб-структурами Интернета, такими как открытые каталоги, обмен блогами и т. д. Сфера применения SKOS не ограничена только тезаурусами и распространяется на другие типы систем организации знаний, такие как системы классификаций/предметных рубрик, таксономии, глоссарии, контролируемые словари и т. д. SKOS определяет классы и свойства, позволяющие представление отношений лексических единиц, существующих в стандартном тезаурусе, но при этом в стандарте установлен концептуально-ориентированный подход к элементам словаря, где элементарные объекты представляют не термины, а понятия-«концепции», представленные терминами. Каждая концепция SKOS определяется как «ресурс RDF» и может характеризоваться 32 «свойствами RDF». В SKOS семантические отношения между понятиями точно соответствуют иерархическим и ассоциативным отношениям, рекомендованным в ИСО 2788. Они принимают форму трех стандартных «свойств»: “skos:broade” – более широкое и “skos:narrowe” – более узкое для иерархических связей и “skos:related” – связанные, для ассоциативных связей между понятиями. Концепции могут объединяться в концептуальные модели, включающие часть или все элементы контролируемого словаря.

При разработке SKOS с самого начала учитывалась возможность совместимости со стандартами ИСО 2788 и ИСО 5964, которые устанавливали базовые правила разработки тезаурусов. Однако эти стандарты были разработаны для традиционных бумажных технологий и не учитывают специфики цифровых носителей информации и электронных сред коммуникаций.

---

[9] Review of RDF Thesaurus Work, IST, SWAD-Europe Thesaurus Activity, Deliverable 8.2. URL: http://www.w3.org/2001/sw/Europe/reports/thes/8.2/draft01.html (дата обращения 12.01.2024).

[10] SKOS Simple Knowledge Organization System Reference // W3C Recommendation 18 August 2009. URL: http://www.w3.org/TR/2009/REC-skos-reference-20090818/ (дата обращения 12.01.2024).

Появление стандарта SKOS послужило толчком к разработке нового международного стандарта ИСО, учитывающего развитие принципов индексирования ресурсов Интернета и гармонизированного со стандартами W3C. Новый международный стандарт ИСО 25964 «Информация и документация. Тезаурусы и взаимодействие с другими словарями», состоящий из двух частей, был разработан рабочей группой, в состав которой вошли члены из 15 стран под председательством BSI (Великобритания) и секретариатом, возглавляемым NISO (США). Со стороны России в качестве эксперта активное участие в проекте принимал известный лингвист В.Н. Белоозеров (ВИНИТИ, ТК 191). Первая часть стандарта ИСО 25964-1 «Тезаурус для информационного поиска»[11] была опубликована в августе 2011 г. Он обновляет, пересматривает и заменяет стандарты ИСО 2788 и ИСО 5964. Структура тезауруса по ИСО 25964-1 была определена как модель данных, изначально разработанная и представленная в национальном стандарте Великобритании BS 8723 «Структурированные словари для информационного поиска. Руководство»[12].

В 2015 г. первая часть ИСО 25964 была введена в российскую систему стандартизации как ГОСТ Р 7.0.91-2015 «СИБИД. Тезаурусы для информационного поиска». Стандарт устанавливает рекомендации для развития и ведения информационно-поисковых тезаурусов. Эти рекомендации можно применить и к другим словарям, применяемым в процессе поиска информации по всем видам информационных ресурсов. Стандарт может быть применен как к одноязычным, так и к многоязычным тезаурусам, хотя в него не вошли разделы, касающиеся построения многоязычных тезаурусов, поскольку в России эти вопросы сегодня регулируются стандартом ГОСТ 7.24-2007[13].

В стандарте представлены определения основных терминов и концепций, используемых при создании ИПС, типы связей между

---

[11] ISO 25964-1:2011 Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval // ISO. URL: https://www.iso.org/standard/53657.html (дата обращения 12.01.2024).

[12] The official development website for BS 8723: Structured Vocabularies for Information Retrieval // Website for BS 8723. URL: https://schemas.bs8723.org/model/ (дата обращения 12.01.2024).

[13] ГОСТ 7.24-2007 Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению. URL: https://docs.cntd.ru/document/1200057506 (дата обращения 12.01.2024).

терминами и концепциями ИПС, рекомендации по созданию и ведению, оформлению и представлению тезауруса, а также руководство по применению фасетного анализа к тезаурусам.

Российский стандарт также строго определяет различия между терминами и концепциями. В нем сохраняются теги BT, NT и RT, поскольку они широко используются в существующих тезаурусах, но даются разъяснения, что отношения, которые они указывают, относятся к понятиям, а не к терминам. «Каждое понятие в тезаурусе представлено одним дескриптором в каждом языке и некоторым количеством аскрипторов. Нотация, примечания и родовидовые отношения применяются к понятию в целом, а не к дескриптору как таковому. Каждому понятию может быть присвоен идентификатор»[14].

Основные принципы, определенные в указанном стандарте, иллюстрируются моделью структуры данных, которая может служить основой для создания тезауруса. Модель данных содержит пять основных классов: тезаурус, массив тезауруса, концепция тезауруса, термин тезауруса и примечание. Атрибуты для каждого класса и ассоциации классов отражают все функции тезаурусов, рекомендуемые в тексте. Модель сопровождается четкими пояснительными примечаниями. Модель представлена графически с использованием Унифицированного Языка Моделирования UML (Unified Modeling Language), который строго определяет структуру и функции тезауруса, что позволяет ее использовать для спецификации по созданию и использованию программного обеспечения ИПС, исключая неопределенности.

Учитывая, что некоторые пользователи могут быть незнакомы со спецификацией UML, стандарт предлагает альтернативное табличное представление модели данных ИПТ. Элементы данных и атрибуты в диаграммах UML и в таблицах данных идентичны по содержанию, хотя табличное представление менее наглядно, и в нем не отражены некоторые детали отношений, связанные с нотацией UML.

Российский стандарт гармонизирован с ИСО 25964-1, но является модифицированным по отношению к нему с учетом российской специфики. В российский стандарт не вошли перечни сокращений терминов на языках, не применяющихся в практике национальной стандартизации, информационные приложения

[14] ГОСТ Р 7.0.91-2015 Система стандартов по информации, библиотечному и издательскому делу. Тезаурусы для информационного поиска. URL: https://docs.cntd.ru/document/1200129056 (дата обращения 12.01.2024).

с примерами интерфейсов тезаурусов и ссылками на сайты иностранных организаций. Приведенные в разделах примеры также представлены на русском языке. Кроме того, из стандарта исключены разделы, регулирующие построение и оформление мультиязычных тезаурусов, так как они не содержат принципиальных положений, отличающихся от положений действующего в России ГОСТ 7.24-2007, определяющего правила для многоязычных информационно-поисковых тезаурусов.

Вторая часть международного стандарта ИСО 25964-2 «Взаимодействие с другими словарями была опубликована в 2013 г.[15] В стандарте описаны методы сопоставления структуры и элементов тезаурусов и других типов словарей, которые обычно используются для поиска информации. Основное внимание уделяется принципам моделирования и практике и сопоставления понятий, представленных в контролируемых словарях различных видов. Стандарт дает рекомендации по созданию и поддержанию соответствий между несколькими тезаурусами или между тезаурусами и другими типами словарей. Область применения включает взаимодействие со схемами классификаций, таксономиями, словарями предметных рубрик, онтологиями, терминосистемами, авторитетными файлами и словарями синонимических рядов. Стандарт содержит указания по разработке сопоставительных таблиц, которые указывают соответствия элементов различных схем и словарей.

В настоящее время в рабочей группе технического комитета по стандартизации ТК 191, включающей экспертов из ГПНТБ России и ВИНИТИ РАН, завершается работа над проектом введения второй части ИСО 25964 в российскую систему стандартизации как ГОСТ Р 7.0.106 «СИБИД. Взаимодействие тезаурусов и других словарей», модифицированного международному. Ввод в действие и публикация нового стандарта планируется в 2024 г.

Концепция Семантической паутины подразумевает достижение функциональной совместимости всего разнообразия стандартов и протоколов, каждый из которых должен обеспечивать общее взаимодействие информационных систем. Достижение такой совместимости возможно только при условии сотрудничества между разработчиками каждого фрагмента, составляющего общее коммуникационное пространство Интернета. Модели данных, представленные в стандартах W3C SKOS и ИСО 25964, в значи-

[15] ISO 25964-2:2013 Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies // ISO. URL: https://www.iso.org/standard/53658.html (дата обращения 12.01.2024).

тельной степени совместимы, что явилось результатом тесного сотрудничества экспертов рабочих групп на стадии разработки. Тем не менее их модели данных не идентичны. Стандарт ИСО 25964 определяет правила для формирования любых тезаурусов, независимо от предметной области их использования, включая веб-приложения, в то время как SKOS ориентирован исключительно на веб-приложения, использующие все виды систем организаций знаний (KOS), в том числе схемы классификации, не включенные в ИСО 25964. Однако, несмотря на различия в областях применения, разработчикам удалось достичь хорошей согласованности между компонентами моделей данных. Кроме того, существует механизм гармонизации различий, возникающих в процессе их развития. Элементы, представленные в модели ИСО 25964 и отсутствующие в базовой модели SKOS (например, «сложная эквивалентность» (compound equivalence)[16], могут быть включены в расширенную модель SKOS-XL, которая дополняется по необходимости.

Заключение

Развитие технических средств Интернета как глобальной коммуникационной среды сопровождается появлением новых видов электронных информационных ресурсов, включаемых в библиотечные фонды и участвующих в информационном обмене и информационном поиске. Разнообразие новых видов ресурсов влечет за собой появление новых технологий их индексирования, которые могут сильно отличаться от традиционных средств и методов индексирования печатных книг. Проблема совместимости разнообразных технологий индексирования информационных ресурсов, представленных в Интернете, является ключевой в реализации концепции Семантической паутины. В этих условиях только стандартизация может обеспечить дальнейшее развитие поисковых сервисов информационных систем библиотек и других организаций информационного профиля. Применение новых стандартов поисковых тезаурусов позволит повысить эффективность представления ресурсов библиотечных электронных каталогов в разнородных поисковых системах и, как следствие, повысит их востребованность среди современных пользователей информационных сервисов Интернета.

[16] ГОСТ Р 7.0.91–2015 Система стандартов по информации, библиотечному и издательскому делу. Тезаурусы для информационного поиска (п. 2.8).

Литература

Антопольский, Вайсберг, Зарувинская 1976 – *Антопольский А.Б., Вайсберг А.М., Зарувинская Л.А.* Принципы создания и функционирования автоматизированной системы ведения информационных языков // Научно-техническая информация. Серия 2. 1976. № 6. С. 8–12.

Данилов, Уманский 1969 – *Данилов М.П., Уманский А.Н.* О принципах строения и путях создания единой системы классификации печатных изданий и документальных материалов. М.: ВНИИКИ, 1969.

Тимошенко 2023 – *Тимошенко И.В.* Стандартизация информационных видов деятельности. Система стандартов по информации, библиотечному и издательскому делу. М.: Знание-М, 2023.

Шемакин 1972 – *Шемакин Ю.И.* Тезаурус научно-технических терминов. М.: Воениздат, 1972.

Thesaurus 1967 – Thesaurus of Engineering and Scientific Terms. A List of Engineering and Related Scientific Terms and Their Relationships for Use as a Vocabulary Reference in Indexing and Retrieving Technical Information. New York, NY: Engineers Joint Council, 1967. 690 p.

Jahns 2012 – Guidelines for Subject Access in National Bibliographies // IFLA Working Group on Guidelines for Subject Access by National Bibliographic Agencies / Y. Jahns (ed.). Berlin: De Gruyter Saur, 2012. 109 p. (IFLA Series on Bibliographic Control, vol. 45).

Petersen 1994 – *Petersen T.* Art and Architecture Thesaurus. In 5 vols. New York: Oxford University Press, 1994.

References

Antopolsky, A.B., Vaisberg, A.M. and Zaruvinskaya, L.A. (1976), “Principles of creation and functioning of an automated system for maintaining information languages”, *Nauchno-tekhnicheskaya informatsiya, series 2*, vol. 6, pp. 8–12.

Danilov, M. and Umanskii, A. (1969), *O principakh stroeniya i putyakh sozdaniya edinoi sistemy klassifikatsii pechatnykh izdanii i dokumentalnykh materialov* [On the principles of structure and ways of creating a unified classification system for printed publications and documentary materials], VNIIKI, Moscow, Russia.

Thesaurus of Engineering and Scientific Terms (1967), Thesaurus of Engineering and Scientific Terms. A List of Engineering and Related Scientific Terms and Their Relationships for Use as a Vocabulary Reference in Indexing and Retrieving Technical Information, Engineers Joint Council, New York, NY, USA.

Jahns, Y. (ed.) (2012), Guidelines for Subject Access in National Bibliographies // IFLA Working Group on Guidelines for Subject Access by National Bibliographic Agencies. De Gruyter Saur, Berlin, Germany. (IFLA Series on Bibliographic Control, vol. 45)

Petersen, T. (1994), Art and Architecture Thesaurus. In 5 vols., Oxford University Press, New York, USA.

Shemakin, Yu.I. (1972), *Tezaurus nauchno-tekhnicheskikh terminov* [Thesaurus of scientific and technical terms], Voenizdat, Moscow, Russia.

Timoshenko, I.V. (2023), *Standartizatsiya informatsionnykh vidov deyatel'nosti. Sistema standartov po informatsii, bibliotechnomu i izdatel'skomu delu* [Standardization of information activities. The system of standards for information, library and publishing], Znaniye-M, Moscow, Russia.

Информация об авторе

Игорь В. Тимошенко, кандидат технических наук, Московский государственный институт культуры, Химки, Моск. обл., Россия; 141406, Химки, Моск. обл., Библиотечная ул., д. 7, корп. 2;

Государственная публичная научно-техническая библиотека России, Москва, Россия; 123298, Москва, Россия, 3-я Хорошевская ул., д. 17; timigor@gpntb.ru

Information about the author

Igor V. Timoshenko, Cand. of Sci. (Technology), Moscow State Institute of Culture, Khimki, Moscow region, Russia; bld. 7/2, Bibliotechnaya Str., Khimki, Moscow region, 14140, Russia;

Russian National Public Library for Science and Technology, Moscow, Russia; bld. 17, Khoroshevskaya 3rd Str, Moscow, 123298, Russi; timigor@gpntb.ru

УДК 681.513:004.4
DOI: 10.28995/2686-679X-2024-2-26-43

Программная реализация модели Co-Active Neuro-Fuzzy Inference System

Андрей П. Титов
Российский технологический университет МИРЭА,
Москва, Россия, titov_and@mail.ru

Аннотация. Статья посвящена реализации нейронной сети с нечеткой логикой на основе модели Co-Active Neuro-Fuzzy Inference System (CANFIS). Модель CANFIS представляет собой адаптивную нейро-нечеткую систему, которая сочетает в себе нейронные сети и нечеткую логику для обработки данных с неопределенностью и нечеткостью. CANFIS использует нечеткие правила и механизмы вывода для преобразования входных данных в выходные значения. Она состоит из нескольких слоев, включая входной слой, скрытые слои и выходной слой, где каждый слой содержит нейроны, выполняющие нечеткую активацию и вывод результатов. Актуальность работы состоит в том, что программная реализация модели CANFIS, основанная на STL языка C++, имеет большое значение в области машинного обучения, искусственного интеллекта и анализа данных. Результаты этой работы могут быть применены в различных областях, в том числе при принятии решений на основе нечеткой логики. Особенность изученной и разработанной модели заключается в создании адаптивной модели, способной моделировать системы с неопределенностью и размытостью. Разработанная модель способна обрабатывать данные и принимать решения на основе нечетких правил. CANFIS находит применение в различных областях, включая прогнозирование, управление, классификацию и анализ данных. Можно сделать вывод о том, что разработанная нейронная сеть с нечеткой логикой может быть эффективно применена в различных областях, где используется прогнозирование временных рядов, системное управление и принятие решений на основе нечеткой информации.

Ключевые слова: нейронные сети, адаптивная нейронная сеть, CANFIS, машинное обучение, нейроны, нечеткая логика, нейро-нечеткие системы

Для цитирования: Титов А.П. Программная реализация модели Co-Active Neuro-Fuzzy Inference System // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 26–43. DOI: 10.28995/2686-679X-2024-2-26-43

Software implementation of the Co-Active Neuro-Fuzzy Inference System

Andrei P. Titov
Russian Technological University MIREA, Moscow, Russia,
titov_and@mail.ru

Abstract. The article deals with the implementation of a neural network with fuzzy logic based on the Co-Active Neuro-Fuzzy Inference System (CANFIS) model. The CANFIS model is an adaptive neuro-fuzzy system that combines neural networks and fuzzy logic for processing data with uncertainty and fuzziness. CANFIS uses fuzzy rules and output mechanisms to convert input data into output values. It consists of several layers, including an input layer, hidden layers and an output layer, where each layer contains neurons performing fuzzy activation and output of results. The relevance of the work lies in the fact that the software implementation of the CANFIS model, based on the STL of the C++ language, is of great importance in the field of machine learning, artificial intelligence and data analysis. The work's results can be applied in various fields, including when making decisions based on fuzzy logic. Special feature of the studied and developed model is to create an adaptive model capable of modeling systems with uncertainty and blurriness. The developed model is able to process data and make decisions based on fuzzy rules. CANFIS finds applications in various fields, including forecasting, management, classification and data analysis. It can be concluded that the developed neural network with fuzzy logic can be effectively applied in various fields where time series forecasting, system management and decision-making based on fuzzy information are used.

Keywords: neural networks, adaptive neural network, CANFIS, machine learning, neurons, fuzzy logic, neuro-fuzzy systems

For citation: Titov, A.P. (2024), "Software implementation of the Co-Active Neuro-Fuzzy Inference System", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 2, pp. 26–43, DOI: 10.28995/2686-679X-2024-2-26-43

Введение

Модель CANFIS (Co-Active Neuro-Fuzzy Inference System) представляет собой гибкую адаптивную сеть, основанную на нечетком выводе, которая применяется для прогнозирования и моделирования сложных систем. Она объединяет преимущества нечеткой логики и нейронных сетей, создавая мощный инструмент для анализа и решения сложных задач.

CANFIS состоит из двух ключевых компонентов: нечеткой сети и адаптивной сети. Нечеткая сеть отвечает за обработку нечетких правил и логического вывода, а адаптивная сеть выполняет обучение и адаптацию модели на основе имеющихся данных.

Нечеткая логика за несколько десятилетий превратилась в мощный инструмент для построения моделей приближенных рассуждений человека в задачах принятия решений в условиях неопределенности, классификации и анализа данных. Математический аппарат теории нечетких множеств позволяет построить модель объекта, основываясь на нечетких рассуждениях и правилах.

Нечеткие модели описывают явления и процессы реального мира на естественном языке при помощи лингвистических переменных, а механизм нечеткого вывода прозрачен и понятен человеку. Эти преимущества обусловили широкое применение нечеткой логики для решения задач автоматического управления, принятия решений, прогнозирования в различных прикладных областях науки, техники и экономики [Титов 2024].

Программная реализация моделей нейронной сети, в том числе и CANFIS

Использование нескольких моделей позволит нам понять правильность работы нашего метода. С этой целью мы произведем программирование нескольких моделей нейронной сети, а именно Гауссовская функция принадлежности, обобщенная функция принадлежности Белл, модель Цукамото, TSK, модель Co-Active Neuro-Fuzzy Inference System.

Полный алгоритм, разработанный программной реализацией, представлен на рис. 1.

В начале представленного алгоритма выполняется подключение необходимых библиотек, изображенных на рис. 2.

Каждая из этих библиотек предоставляет определенный набор функций и классов, которые можно использовать в программе. Использование этих библиотек позволяет программистам легко использовать готовые решения и функциональности, упрощает написание кода и повышает производительность разработки.

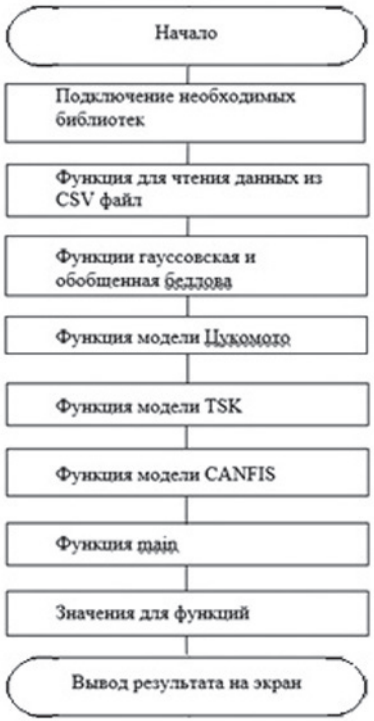


Рис. 1. Схема алгоритма предложенной модели Co-Active Neuro-Fuzzy Inference System

```
#include <iostream>
#include <vector>
#include <string>
#include <fstream>
#include <sstream>
#include <cmath>
#include <algorithm>

using namespace std;
```

Рис. 2. Библиотеки

Далее идет функция для чтения данных из CSV-файла и сохранения их в векторе, которая изображена на рис. 3.

```
// Функция для чтения данных из CSV файла и сохранения их в векторе
vector<vector<double>> readCSVData(const string& filename) {
    vector<vector<double>> data;
    ifstream file(filename);
    if (!file) {
        cout << "Ошибка при открытии файла: " << filename << endl;
        return data;
    }
    string line;
    while (getline(file, line)) {
        stringstream ss(line);
        string cell;
        vector<double> rowData;

        while (getline(ss, cell, ',')) {
            double value;
            stringstream converter(cell);
            converter >> value;
            rowData.push_back(value);
        }
        data.push_back(rowData);
    }
    return data;
}
```

Рис. 3. Функция для чтения данных из CSV-файл

Далее создаются функции гауссовская и обобщенная Беллова, изображенные на рис. 4.

```
// Функция принадлежности - гауссовская функция
double gaussianMembership(double x, double mean, double sigma) {
    return exp(-pow(x - mean, 2) / (2 * pow(sigma, 2)));
}

// Функция принадлежности - обобщенная беллова функция
double generalizedBellMembership(double x, double a, double b, double c) {
    return 1 / (1 + pow(abs((x - c) / a), 2 * b));
}
```

Рис. 4. Функции гауссовская и обобщенная Беллова

Пояснение каждой из функций:

1. `double gaussianMembership(double x, double mean, double sigma)`: Эта функция реализует гауссовскую функцию принадлежности. Гауссовская функция имеет колоколообразную форму и широко используется в нечетких системах. Она принимает три параметра: `x` (значение элемента), `mean` (среднее значение) и `sigma` (стандартное отклонение). Функция вычисляет степень принадлежности элемента `x` к множеству на основе гауссовского распределения.

2. `double generalizedBellMembership(double x, double a, double b, double c)`: Эта функция реализует обобщенную беллову функцию принадлежности. Обобщенная беллова функция имеет форму, которая может быть изменена, чтобы адаптироваться к различным видам данных. Она принимает четыре параметра: `x` (значение элемента), `a`, `b` и `c`. Параметры `a`, `b` и `c` определяют форму функции. Функция вычисляет степень принадлежности элемента `x` к множеству на основе обобщенной белловой функции.

Обе функции используют математические операции, такие как возведение в степень (`pow`), вычисление абсолютного значения (`abs`) и экспоненту (`exp`), для расчета степени принадлежности. Эти функции могут быть полезны при моделировании нечетких систем, где необходимо определить степень принадлежности элементов к различным множествам.

Создаем функцию модели Цукомото, изображенную на рис. 5.

```
// Модель нечеткой системы – модель Цукамото
double tsukamotoModel(const vector<double>& input, const
vector<vector<double>>& rules, const vector<double>& output) {
    double numerator = 0.0;
    double denominator = 0.0;
    for (size_t i = 0; i < rules.size(); i++) {
        double ruleStrength = 1.0;
        // Рассчитываем силу правила
        for (size_t j = 0; j < input.size(); j++) {
            ruleStrength = min(ruleStrength, rules[i][j]);
        }
        // Умножаем силу правила на выходное значение
        double weightedOutput = ruleStrength * output[i];
        numerator += weightedOutput;
        denominator += ruleStrength;
    }
    if (denominator != 0.0) {
        return numerator / denominator;
    }
    else {
        return 0.0;
    }
}
```

Рис. 5. Функция модели Цукомото

Этот отрывок кода реализует модель нечеткой системы, известную как модель Цукамото. Модель Цукамото используется для принятия решений на основе нечетких правил, которые описывают взаимодействие между входными и выходными переменными.

Объяснение работы этой модели:

1. `double tsukamotoModel(const vector<double>& input, const vector<vector<double>>& rules, const vector<double>& output)`: Эта функция принимает три параметра: `input` (вектор входных переменных), `rules` (матрица правил) и `output` (вектор выходных переменных). Функция вычисляет выходное значение модели на основе входных переменных и правил.

2. `double numerator = 0.0;` и `double denominator = 0.0;`: Переменные `numerator` и `denominator` используются для накопления значений при вычислении выходного значения модели.

3. `for (size_t i = 0; i < rules.size(); i++) { ... }`: Этот цикл перебирает все правила в матрице `rules`.

4. `double ruleStrength = 1.0;`: Переменная `ruleStrength` инициализируется значением 1.0 и представляет силу текущего правила.

5. Внутренний цикл `for (size_t j = 0; j < input.size(); j++) { ... }` рассчитывает силу правила путем нахождения минимального значения между `ruleStrength` и соответствующим элементом вектора входных переменных `input`. Это основано на предположении, что силу правила определяет наименьшее значение среди всех входных переменных.

6. `double weightedOutput = ruleStrength * output[i];`: Вычисляется взвешенное значение выходной переменной путем умножения силы правила `ruleStrength` на соответствующий элемент вектора выходных переменных `output`.

7. `numerator += weightedOutput;` и `denominator + = rule Strength;`: Накапливаются значения для дальнейших вычислений.

8. В конце цикла проверяется, является ли `denominator` равным нулю. Если это так, то функция возвращает 0.0. В противном случае вычисляется отношение `numerator / denominator` и возвращается в качестве выходного значения модели.

Таким образом, данный код выполняет вычисление выходного значения модели Цукамото на основе входных переменных, матрицы правил и выходных переменных.

Далее создается функция модели TSK, изображенная на рис. 6.

Данный код реализует модель нечеткой системы, известную как модель TSK (Takagi-Sugeno-Kang). Модель TSK является одним из вариантов моделей нечетких систем и используется для принятия решений на основе нечетких правил.

```
// Модель нечеткой системы – модель TSK (Takagi-Sugeno-Kang)
double tskModel(const vector<double>& input, const vector<vector<double>>&
rules, const vector<double>& output) {
    double result1 = 0.0;

    for (size_t i = 0; i < rules.size(); i++) {
        double ruleStrength = 1.0;

        // Рассчитываем силу правила
        for (size_t j = 0; j < input.size(); j++) {
            ruleStrength = min(ruleStrength, rules[i][j]);
        }

        // Умножаем силу правила на выходное значение и суммируем
        result1 += ruleStrength * output[i];
    }

    return result1;
}
```

Рис. 6. Функция модели TSK

Объяснение работы модели:

1. `double tskModel(const vector<double>& input, const vector<vector<double>>& rules, const vector<double>& output)`: Эта функция принимает три параметра: `input` (вектор входных переменных), `rules` (матрица правил) и `output` (вектор выходных переменных). Функция вычисляет выходное значение модели TSK на основе входных переменных и правил.

2. `double result1 = 0.0;`: Переменная `result1` инициализируется значением 0.0 и представляет суммарное выходное значение модели.

3. `for (size_t i = 0; i < rules.size(); i++) { ... }`: Этот цикл перебирает все правила в матрице `rules`.

4. `double ruleStrength = 1.0;`: Переменная `ruleStrength` инициализируется значением 1.0 и представляет силу текущего правила.

5. Внутренний цикл `for (size_t j = 0; j < input.size(); j++) { ... }` рассчитывает силу правила путем нахождения минимального значения между `ruleStrength` и соответствующим элементом вектора входных переменных `input`. Это основано на предположении, что силу правила определяет наименьшее значение среди всех входных переменных.

6. `result1 += ruleStrength * output[i];`: Вычисляется взвешенное значение выходной переменной путем умножения силы правила `ruleStrength` на соответствующий элемент вектора выходных переменных `output`, а затем результат суммируется с `result1`.

7. В конце цикла возвращается значение `result1` в качестве выходного значения модели TSK.

Таким образом, данный код выполняет вычисление выходного значения модели TSK на основе входных переменных, матрицы правил и выходных переменных.

Создание функции модели CANFIS изображено на рис. 7.

Данный код реализует модель нечеткой системы, известную как модель CANFIS (Co-Active Neuro-Fuzzy Inference System). Модель CANFIS представляет собой комбинацию нейронных сетей и нечеткой логики и используется для прогнозирования и аппроксимации данных.

Рассмотрим более детально программный код этой модели:

1. `double canfisModel(const vector<double>& input, const vector<vector<double>>& mainLayer, const vector<vector<double>>& hiddenLayers, const vector<double>& output)`: Эта функция принимает четыре параметра: `input` (вектор входных переменных), `mainLayer` (матрица весов основного слоя), `hiddenLayers` (матрица весов скрытых слоев) и `output` (вектор выходных переменных). Функция вычисляет выходное значение модели CANFIS на основе входных переменных и весовых матриц.

2. `vector<double> hiddenOutputs(hiddenLayers.size());`: Создается вектор `hiddenOutputs`, который будет хранить выходы скрытых слоев.

3. Цикл `for (size_t i = 0; i < hiddenLayers.size(); i++) { ... }` вычисляет выходы скрытых слоев.

4. Внутренний цикл `for (size_t j = 0; j < hiddenLayers[i].size(); j++) { ... }` вычисляет сумму взвешенных входов для каждого нейрона в текущем скрытом слое. Сумма взвешенных входов вычисляется путем умножения весов из `hiddenLayers` на соответствующие элементы из вектора входных переменных `input`.

5. `hiddenOutputs[i] = gaussianMembership(sum, 0.0, 1.0);`: Выход скрытого слоя вычисляется путем применения функции принадлежности (гауссовской функции) к сумме взвешенных входов.

```
// Модель нечеткой системы – модель CANFIS (Co-Active Neuro-Fuzzy Inference System)
double canfisModel(const vector<double>& input, const vector<vector<double>>&
mainLayer, const vector<vector<double>>& hiddenLayers, const vector<double>& output) {
    vector<double> hiddenOutputs(hiddenLayers.size());
    // Вычисляем выходы скрытых слоев
    for (size_t i = 0; i < hiddenLayers.size(); i++) {
        double sum = 0.0;
        // Вычисляем сумму взвешенных входов для каждого нейрона в скрытом слое
        for (size_t j = 0; j < hiddenLayers[i].size(); j++) {
            sum += hiddenLayers[i][j] * input[j];
        }
        // Применяем функцию принадлежности к сумме взвешенных входов
        hiddenOutputs[i] = gaussianMembership(sum, 0.0, 1.0);
    }
    // Вычисляем выход основного слоя
    double mainOutput = 0.0;
    for (size_t i = 0; i < mainLayer.size(); i++) {
        double sum = 0.0;
        // Вычисляем сумму взвешенных выходов скрытых слоев
        for (size_t j = 0; j < hiddenOutputs.size(); j++) {
            sum += mainLayer[i][j] * hiddenOutputs[j];
        }
        // Применяем функцию принадлежности к сумме взвешенных выходов
        mainOutput += generalizedBellMembership(sum, 1.0, 1.0, 1.0);
    }
    // Вычисляем итоговый выход нейронной сети
    double result = 0.0;
    for (size_t i = 0; i < output.size(); i++) {
        result += output[i] * mainOutput;
    }
    return result;
}
```

Рис. 7. Функция модели CANFIS

6. После вычисления выходов скрытых слоев происходит вычисление выхода основного слоя.

7. Цикл `for (size_t i = 0; i < mainLayer.size(); i++) { ... }` вычисляет сумму взвешенных выходов скрытых слоев для каждого нейрона в основном слое.

8. `mainOutput += generalizedBellMembership(sum, 1.0, 1.0, 1.0);`: Выход основного слоя вычисляется путем применения функции принадлежности (обобщенной белловой функции) к сумме взвешенных выходов скрытых слоев.

9. Затем происходит вычисление итогового выхода нейронной сети путем умножения выхода основного слоя на соответствующие элементы из вектора выходных переменных `output` и их суммирования.

10. В конце функции возвращается результат вычисления итогового выхода нейронной сети.

Таким образом, данный код выполняет прямой проход (forward pass) модели CANFIS, где входные переменные проходят через скрытые слои и основной слой, и вычисляется итоговый выход нейронной сети на основе весовых матриц и функций принадлежности.

Создадим основную функцию main, в ней делаем запрос на имя файла, который будет читать программа. Данные для обучения и тестирования изображены на рис. 8.

Представленный код выполняет чтение данных из файлов для обучения и тестирования нейронной сети, а затем выводит данные для тестирования на экран.

1. `string trainingDataFile;`: Объявляется переменная `trainingDataFile`, которая будет содержать имя файла с данными для обучения нейронной сети.

2. `cout << «Введите имя файла с данными для обучения»;`: Эта строка выводит сообщение пользователю, запрашивая имя файла с данными для обучения.

3. `cin >> trainingDataFile;`: С помощью оператора ввода `>>` пользователь вводит имя файла с данными для обучения, которое сохраняется в переменной `trainingDataFile`.

4. `vector<vector<double>> trainingData = readCSVData(trainingDataFile);`: Вызывается функция `readCSVData`, которая принимает имя файла и возвращает двумерный вектор (`vector<vector<double>>`) с данными из файла. Эти данные присваиваются переменной `trainingData`, которая будет использоваться для обучения нейронной сети.

5. Аналогично происходит чтение и сохранение данных из файла для тестирования нейронной сети. Пользователю также выводится сообщение для ввода имени файла с данными для тестирования.

6. `cout << «Данные для тестирования» << endl;`: Выводится сообщение, указывающее на то, что следующие данные предназначены для тестирования.

```
int main() {
    setlocale(LC_ALL, "Russian");
    //Чтение файла с данными для обучения нейронной сети
    string trainingDataFile;
    cout << "Введите имя файла с данными для обучения: ";
    cin >> trainingDataFile;
    vector<vector<double>> trainingData = readCSVData(trainingDataFile);
    // Чтение файла с данными для тестирования нейронной сети
    string testingDataFile;
    cout << "Введите имя файла с данными для тестирования: ";
    cin >> testingDataFile;
    vector<vector<double>> testingData = readCSVData(testingDataFile);
    // Вывод данных
    cout << "Данные для тестирования: " << endl;
    for (const auto& row : testingData) {
        for (const auto& value : row) {
            cout << value << " ";
        }
        cout << endl;
    }
```

Рис. 8. Функция main

7. Цикл `for (const auto& row : testingData) { ... }` перебирает каждую строку вектора `testingData`, содержащую данные для тестирования.

8. Внутренний цикл `for (const auto& value : row) { ... }` перебирает каждое значение в текущей строке и выводит его на экран с использованием оператора `cout`.

Исходный код позволяет пользователю ввести имена файлов с данными для обучения и тестирования нейронной сети, читает данные из этих файлов и выводит данные для тестирования на экран. Это часто используется при подготовке данных для обучения и проверки работы нейронной сети.

Загружаем значения из CSV-файла и используем их для функции CANFIS, как изображено на рис. 9 и 10.

```
        // Пример использования функции canfisModel
        if (testingData.size() > 0) {
            vector<double> input(testingData[0].begin(), testingData[0].begin() + 3); // Пример:
первые 3 значения в первой строке - input
            vector<vector<double>> mainLayer(3, vector<double>(2)); // Пример: следующие 6
значений – mainLayer
            vector<vector<double>> hiddenLayers(2, vector<double>(2)); // Пример: следующие 4
значения – hiddenLayers
            vector<double> output(testingData[0].begin() + 13, testingData[0].begin() + 15); //
Пример: следующие 2 значения - output
            // Заполняем значения для mainLayer из testingData
            size_t dataIndex = 3;
            for (size_t i = 0; i < mainLayer.size(); i++) {
                for (size_t j = 0; j < mainLayer[i].size(); j++) {
                    mainLayer[i][j] = testingData[0][dataIndex];
                    dataIndex++;
                }
            }
            // Заполняем значения для hiddenLayers из testingData
            size_t dataInd = 9;
            for (size_t i = 0; i < hiddenLayers.size(); i++) {
                for (size_t j = 0; j < hiddenLayers[i].size(); j++) {
                    hiddenLayers[i][j] = testingData[0][dataInd];
                    dataInd++;
                }
            }
              // Вычисление результата с использованием модели CANFIS
            double result = canfisModel(input, mainLayer, hiddenLayers, output);
            // Вывод результата
            cout << "Результат модели CANFIS : " << result << endl;
        }
        return 0;
    }
```

Рис. 9. Загрузка значений из файла CSV в переменные

```
Введите имя файла с данными для обучения: trainingdata.csv
Введите имя файла с данными для тестирования: testdata.csv
Данные для тестирования:
1 2 3 0.5 0.8 0.7 0.9 0.4 0.6 0.3 0.6 0.2 0.4 0.2 0.5
Результат гауссовской функции принадлежности: 0.0439369
Результат обобщенной белловой функции принадлежности: 0.941176
Результат модели Цукамото: 0.475
Результат модели TSK: 0.38
Результат модели CANFIS : 1.8456
```

Рис. 10. Вывод результата на экран по функции CANFIS

Детально рассмотрим работу этого кода с использованием функции `canfisModel` для вычисления результата с использованием модели CANFIS (Co-Active Neuro-Fuzzy Inference System).

1. `if (testingData.size() > 0) {`: Проверяется, что `testingData` содержит данные (непустой вектор).

2. `vector<double> input(testingData[0].begin(), testingData[0].begin() + 3);`: Создается вектор `input`, который содержит первые 3 значения из первой строки `testingData`. Эти значения используются в качестве входных данных для модели CANFIS.

3. `vector<vector<double>> mainLayer(3, vector<double>(2));`: Создается двумерный вектор `mainLayer` размером 3 × 2. Значения `mainLayer` заполняются из `testingData` начиная с позиции 3. Каждый элемент `mainLayer` используется в модели CANFIS.

4. `vector<vector<double>> hiddenLayers(2, vector<double>(2));`: Создается двумерный вектор `hiddenLayers` размером 2 × 2. Значения `hiddenLayers` заполняются из `testingData` начиная с позиции 9. Каждый элемент `hiddenLayers` используется в модели CANFIS.

5. `vector<double> output(testingData[0].begin() + 13, testingData[0].begin() + 15);`: Создается вектор `output`, который содержит следующие 2 значения из первой строки `testingData`. Эти значения используются в качестве выходных данных для модели CANFIS.

6. Заполняются значения для `mainLayer` и `hiddenLayers` из `testingData`:
   - Сначала устанавливается индекс `dataIndex`, равный 3.
   - Затем значения из `testingData` начиная с позиции 3 заполняются в `mainLayer` построчно.
   - После этого устанавливается индекс `dataInd`, равный 9.
   - Значения из `testingData` начиная с позиции 9 заполняются в `hiddenLayers` построчно.

7. `double result = canfisModel(input, mainLayer, hiddenLayers, output);`: Вызывается функция `canfisModel` с параметрами `input`,

`mainLayer`, `hiddenLayers` и `output`, чтобы вычислить результат модели CANFIS для заданных входных и выходных данных. Результат сохраняется в переменной `result`.

8. `cout << «Результат модели CANFIS» << result << endl;`: Выводится результат модели CANFIS на экран.

Таким образом, данный код демонстрирует использование функции `canfisModel` для вычисления результата с использованием модели CANFIS на основе данных из `testingData`. Результат выводится на экран для дальнейшего анализа или использования.

Заключение

Результатом работы стала программная реализация модели CANFIS (Co-Active Neuro-Fuzzy Inference System) на основе STL языка C++. Проведена обзорная аналитика существующих методов и подходов к моделированию и прогнозированию с использованием нейро-нечетких систем.

В ходе разработки были расширены знания в области нейро-нечетких систем, включая изучение основных принципов функционирования данных моделей, анализ существующих методов и алгоритмов, а также практическую реализацию модели CANFIS с использованием языка C++ и STL. Также были усовершенствованы уже имеющиеся навыки написания кода на языке C++.

Одним из главных преимуществ разработанной нейронной сети является ее универсальность, которая достигнута за счет использования разных моделей. Это позволяет легко модифицировать и адаптировать ее под различные задачи. Кроме того, использование STL языка C++ позволило значительно упростить код и повысить его читабельность.

Таким образом, можно сделать вывод о том, что разработанная нейронная сеть с нечеткой логикой может быть эффективно применена в различных областях, где используется прогнозирование временных рядов, системное управление и принятие решений на основе нечеткой информации.

Литература

Бессмертный 2023 – *Бессмертный И.А.* Системы искусственного интеллекта: учеб. пособие для среднего профессионального образования. М.: Юрайт, 2023. 157 с.

Борисов, Алексеев, Крумберг 1982 – *Борисов А.Н., Алексеев А.В., Крумберг О.А.* Модели принятия решений на основе лингвистической переменной. Рига: Зинатне, 1982. 256 с.

Борисов, Крумберг, Федоров 1990 – *Борисов А.Н., Крумберг О.А., Федоров И.П.* Принятие решений на основе нечетких моделей: примеры использования. Рига: Зинатне, 1990. 184 с.

Воронов, Пименов, Небаев 2023 – *Воронов М.В., Пименов В.И., Небаев И.А.* Системы искусственного интеллекта: учебник и практикум для вузов. М.: Юрайт, 2023. 268 с.

Горбаченко, Ахметов, Кузнецова 2023 – *Горбаченко В.И., Ахметов Б.С., Кузнецова О.Ю.* Интеллектуальные системы: нечеткие системы и сети: Учеб. пособие для вузов. М.: Юрайт, 2023. 105 с.

Колесникова 2023 – *Колесникова С.М.* Когнитивная лингвистика: Учебник для вузов. М.: Юрайт, 2023. 192 с.

Кравченко 2002 – *Кравченко Ю.А.* Перспективы развития гибридных интеллектуальных систем // Перспективные информационные технологии и интеллектуальные системы. 2002. № 3. С. 34–38.

Новиков 2023 – *Новиков Ф.А.* Символический искусственный интеллект: математические основы представления знаний: Учеб. пособие для вузов. М.: Юрайт, 2023. 278 с.

Паклин 2003 – *Паклин Н.Б.* Адаптивные системы нечеткого логического вывода и их приложения // Интеллектуальные системы в производстве. 2003. № 2. С. 138–151.

Поспелов 1986 – *Поспелов Д.А.* Нечеткие множества в моделях управления и искусственного интеллекта. М.: Наука, 1986. 311 с.

Титов 2024 – *Титов А.П.* Анализ моделей адаптивных нейро-нечетких систем // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 1. С. 21–35.

References

Bessmertnyi, I.A. (2023), *Sistemy iskusstvennogo intellekta: ucheb. posobie dlya srednego professional'nogo obrazovaniya* [Artificial intelligence systems. A study guide for secondary vocational education], Yurait, Moscow, Russia.

Borisov, A.N., Alekseev, A.V. and Krumberg, O.A. (1982), *Modeli prinyatiya reshenii na osnove lingvisticheskoi peremennoi* [Models of decision making based on linguistic variables], Zinatne, Riga, Latvia.

Borisov, A.N., Krumberg, O.A. and Fedorov, I.P. (1990), *Prinyatie reshenii na osnove nechetkikh modelei: primery ispol'zovaniya* [Decision making based on fuzzy models. Examples of use], Zinatne, Riga, Latvia.

Gorbachenko, V.I., Akhmetov, B.S. and Kuznetsova, O.Yu. (2023), *Intellektual'nye sistemy: nechetkie sistemy i seti: ucheb. posob. dlya vuzov* [Intelligent systems. Fuzzy systems and networks. A study guide for universities], Yurait, Moscow, Russia.

Kolesnikova, S.M. (2023), *Kognitivnaya lingvistika: uchebnik dlya vuzov* [Cognitive linguistics. A textbook for universities], Yurait, Moscow, Russia.

Kravchenko, Yu.A. (2002), "Prospects for the development of hybrid intelligent systems", *Perspective information technologies and intelligent systems*, no. 3, pp. 34–38.

Novikov, F.A. (2023), *Simvolicheskii iskusstvennyi intellekt: matematicheskie osnovy predstavleniya znanii: ucheb. posob. dlya vuzov* [Symbolic artificial intelligence: mathematical foundations of knowledge representation. A study guide for universities], Yurait, Moscow, Russia.

Paklin, N.B. (2003), "Adaptive systems of fuzzy logical inference and their applications", *Intelligent systems in production*, no. 2, pp. 138–151.

Pospelov, D.A. (1986), *Nechetkie mnozhestva v modelyakh upravleniya i iskusstvennogo intellekta* [Fuzzy sets in control and artificial intelligence models], Nauka, Moscow, Russia.

Titov, A.P. (2024) "Analysis of models of adaptive Neuro-fuzzy Systems", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 1, pp. 21–35.

Voronov M.V., Pimenov V.I., Nebaev I.A. (2023), *Sistemy iskusstvennogo intellekta: uchebnik i praktikum dlya vuzov* [Artificial intelligence systems. A textbook and workshop for universities], Yurait, Moscow, Russia.

Информация об авторе

Андрей П. Титов, кандидат технических наук, доцент, Российский технологический университет МИРЭА, Москва, Россия; 107076, Россия, Москва, ул. Стромынка, д. 20; titov_and@mail.ru

Information about the author

Andrei P. Titov, Cand. of Sci. (Computer Science), associate professor, Russian Technological University MIREA, Moscow, Russia; bld. 20, Stromynka Str., Moscow, 107076, Russia; titov_and@mail.ru

УДК 004.05
DOI: 10.28995/2686-679X-2024-2-44-57

Анализ и оптимизация существующей системы обмена данными реализации продукции с применением функционала программного комплекса “ST Chicago”

Елена О. Шершнева
Сибирский государственный автомобильно-дорожный университет, Омск, Россия, helen_volf@mail.ru

Юрий Ю. Карючин
Сибирский государственный автомобильно-дорожный университет, Омск, Россия, yurykaruchin@gmail.com

Аннотация. В ходе написания статьи изучен объект автоматизации, рассмотрены и модернизированы бизнес-процессы по взаимодействию с автоматизированными системами по управлению автоматизированной дистрибуцией, проведено описание процесса «Как есть» и переработка в процесс «Как надо». Выполнен анализ существующих решений по оптимизации процесса обмена данными и расчета итогов реализации производимых товаров. Созданы на языке структурированных запросов (SQL) хранимые процедуры для построения отчетов при помощи MS Excel и программного комплекса “ST Chicago”.

С точки зрения автоматизированной отчетности был сформирован отдельный отчет, для которого будет производиться отображение фактических продаж по акциям, заведенным в системе «1С: Предприятие» и “ST Chicago”. В процессе разработки отчета были выполнены следующие задачи:

- разработан механизм привязки к номеру онлайн-заказа для определения скидки по товарно-транспортной накладной;
- создан механизм исключения дублей фактических продаж;
- определен механизм проставления скидки на моменте формирование заказа, без использования посткомпенсации;
- выполнена приоритизация акции системы “ST Chicago”, при условии создания одинаковых акций в доступных системах;
- определено сопоставление данных для работы с серверными частями “SAP ERP”, “Power BI”, а также иных версий автоматизированных отчетов.

В результате анализа признаков и классификаторов технологий проектирования информационных систем был выбран канонический класс технологии проектирования.

Ключевые слова: проектирование, аналитика данных, управление автоматизированной дистрибуцией, CRM-системы, процедура

Для цитирования: Шершнева Е.О., Карючин Ю.Ю. Анализ и оптимизация существующей системы обмена данными реализации продукции с применением функционала программного комплекса “ST Chicago” // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 44–57. DOI: 10.28995/2686-679X-2024-2-44-57

Analysis and optimization of the reverse product data exchange system using the "ST Chicago" functional software package

Elena O. Shershneva
The Siberian State Automobile and Highway University, Omsk, Russia, helen_volf@mail.ru

Yurii Yu. Karyuchin
The Siberian State Automobile and Highway University, Omsk, Russia, yurykaruchin@gmail.com

Annotation. In the course of writing the article, the object of automation was studied, business processes for interaction with automated systems for managing automated distribution were reviewed and modernized. The “As Is” process was described and processed into the “As It Should” process. The authors carried out an analysis of existing solutions to optimize the process of data exchange and calculation of the results of the sale of manufactured goods. Stored procedures were created in Structured Query Language (SQL) to generate reports using MS Excel and the ST Chicago software package.

From the point of view of automated reporting, a separate report was generated, for which actual sales for shares entered in the 1C Enterprise and "ST Chicago" systems will be displayed.

During the development of the report, the following tasks were completed:

– a mechanism has been developed to link to the online order number to determine the discount on the consignment note;
– a mechanism has been created to eliminate duplicates of actual sales;
– a mechanism has been defined for applying a discount at the time of order formation, without the use of post-compensation;

– the promotion of the "ST Chicago" system was prioritized, subject to the creation of identical shares in the available systems;
– data comparison has been defined for working with the server parts of SAP ERP, Power BI, as well as other versions of automated reports.

As a result of the analysis of features and classifiers of information system design technologies, the canonical class of design technology was selected.

Keywords: design, data analytics, automated distribution management, CRM systems, procedure

For citation: Shershneva, E.O. and Karyuchin, Yu.Yu. (2024), "Analysis and optimization of the reverse product data exchange system using the "ST Chicago" functional software package", *RSUH/RGGU Bulletin. "Information Sience. Informaton Security. Mathematics" Series,* no. 2, pp. 44–57, DOI: 10.28995/2686-679X-2024-2-44-57

Торговый маркетинг с применением информационных технологий определяется как направление, предполагающее активацию сбыта и продвижения компании за счет увеличения эффективности взаимодействия всех субъектов дистрибутивных точек, с использованием разнообразных программных продуктов. При помощи таких информационных систем производится не только стимулирование сбыта продукции, но и описание показателей качества за счет анализа и систематизации разнообразного количества информации. Применение разнообразных программных продуктов для ведения учета коммерческой деятельности предприятия предполагает постоянную оптимизацию исходных бизнес-процессов за счет интеграции новых методов и информационных технологий для более быстрого построения развернутой аналитики больших данных.

В настоящее время аналогичная форма общения между персональными компьютерами, установленными в торговых точках, и основным сервером предприятия, базирующаяся на внесении новых данных в таблицы реляционных СУБД, реализована практически во всех коммерческих организациях. Эта форма обмена данными предполагает высокую эффективность в связке с применением систем по управлению автоматизированной дистрибуцией.

Исходя из всего описанного выше были поставлены цель и задачи исследования.

Цель: изучение, модернизация и оптимизация бизнес-процессов ведения трейд-маркетинговых акционных мероприятий с применением функционала программного комплекса "ST Chicago".

Задачи:
- изучение объекта модернизации, а именно существующей системы ведения торговых акций;
- рассмотрение и модернизация бизнес-процессов по взаимодействию с автоматизированными системами по управлению автоматизированной дистрибуцией;
- постановка и анализ требований по оптимизации процесса обмена данными и расчета итогов реализации производимых товаров;
- создание на языке структурированных запросов (SQL) хранимых процедур для построения отчетов при помощи MS Excel и программного комплекса “ST Chicago”[1].

Процесс обработки трейд-маркетинговых мероприятий, а также процесс контроля проданной продукции представляет собой широкую совокупность методов, средств, технологий и программных продуктов, необходимых для реализации систем, для осуществления передачи и управления массивами данных без непосредственного использования ручного труда сотрудников организации. То есть определение состава реализационных действий, факта участия в трейд-маркетинговых мероприятиях, канала реализации продукции, а также сбор и передачу информации о реализации внутриэлектронных документов.

Основой автоматизации процесса движения электронных документов является построение архитектуры на сервере организации, содержащем OLTP (Online Transaction Processing) систему, предметно-ориентированную информационную базу данных, а также систему справочников, необходимую для функционирования системы транзакций и агрегаций для передачи в DWH таблицы (Data Warehouse – единое корпоративное хранилище архивных данных из разных источников).

В качестве основной рассматриваемой функции описываемой предметной области выступает формирование автоматизированной отчетности ТМА при помощи системы транзакций и дальнейшее обновление шаблонов MS Excel при помощи специализированных сценариев, сохраненных на сервере организации.

Основными инициаторами описываемого процесса являются автоматизированные дистрибьюторы, передающие информацию о реализации продукции компании для обработки и хранения внутри серверов организации, а также для дальнейшего анализа массива

[1] "ST Chicago" – инструмент для управления бизнес-процессами дистрибуции // ST Systech. URL: https://sys4tec.com/products/st-chicago/ (дата обращения 12.05.2023).

данных, необходимого для определения продаж и планирования процесса ведения бизнеса.

Все вышеперечисленные активности происходят на основании существующих и используемых локальных нормативных актов, приказов, справочников и регламентов. Информационная модель процесса формирования автоматизированной отчетности ТМА в нотации IDEF0 представлена на рис. 1 [Козлов, Лекае, Шаповалова 2019].

Для рассматриваемой предметной области необходимо определить:

– существующую бизнес-архитектуру передачи электронных документов;
– применение системы сопоставления информации внутри справочников с получаемой информацией от контрагентов;
– мероприятия по репликации фактических документов, содержащих информацию о реализации товаров [Борзов 2020].

Рис. 1. Информационная модель процесса формирования автоматизированной отчетности ТМА в нотации IDEF0

На рис. 2 отображена функциональная модель автоматического обновления отчетности программного комплекса “ST Chicago”.

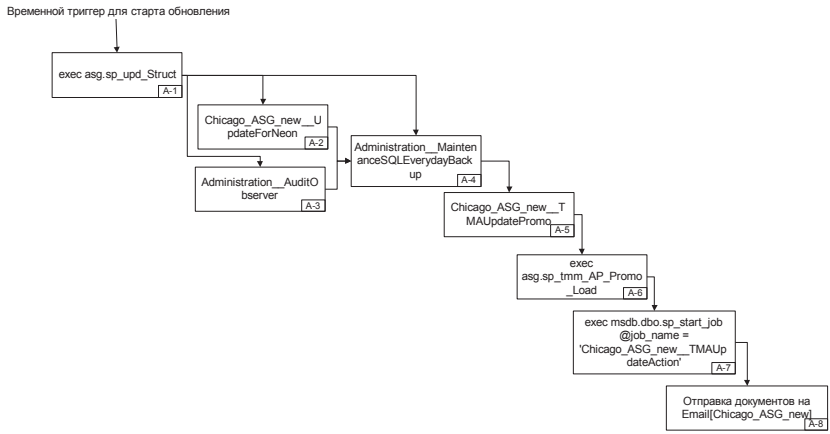


Рис. 2. Функциональная модель автоматического обновления отчетности программного комплекса “ST Chicago”.

В рамках развертывания информационной системы “ST Chicago” используется централизованная архитектура. При использовании данной технологии база данных, СУБД и прикладная программа (приложение) располагаются на одном компьютере (мэйнфрейме или персональном компьютере). Для такого способа организации не требуется поддержки сети и все сводится к автономной работе [Карючин, Мингазов, Пестова 2022].

Работа рассматриваемой информационной системы построена следующим образом:

– база данных в виде набора файлов находится в центре обработки данных “ST Chicago”;
– центр обработки данных включает в себя СУБД и приложение для работы с базой данных;
– все обращения к БД идут через СУБД, которая инкапсулирует внутри себя все сведения о физической структуре БД;
– СУБД инициирует обращения к данным, обеспечивая выполнение запросов пользователя (осуществляя необходимые операции над данными);
– СУБД выполняет по системе транзакций заполнение автоматизированных отчетов и возвращает заполненные двумерные таблицы с подключением к файлам MS Excel для отображения отечности за необходимый период;

– результат СУБД возвращает в приложение, пользовательский интерфейс оператора или агента отображает результат выполнения запросов [Карючин, Мингазов, Пестова 2022].

На рис. 3 отображена схема централизованной архитектуры программного комплекса “ST Chicago”.

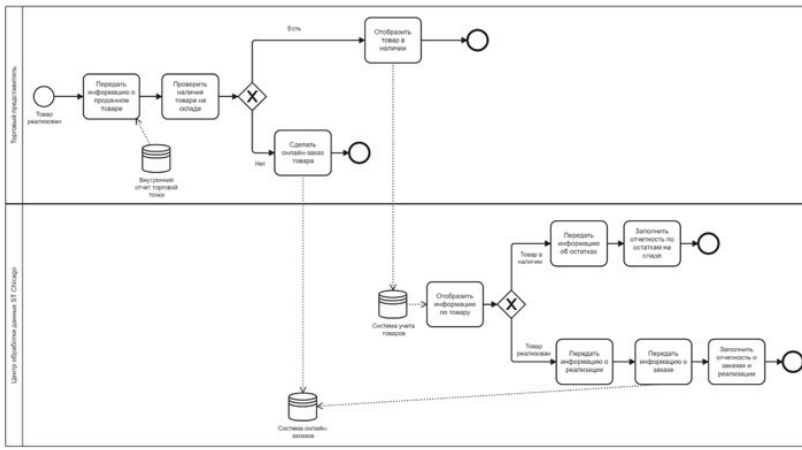


Рис. 3. Централизованная архитектура программного комплекса “ST Chicago”

К основным требованиям, предъявляемым к выбору технологии проектирования модернизированной архитектуры ТМА на базе программного комплекса “ST Chicago”, можно отнести следующие:

– разрабатываемая версия программного продукта должна в полной мере соответствовать требованиям заказчика в рамках возможностей выбранного пути модернизации;
– технология проектирования должна максимально подробно и прозрачно отражать все этапы жизненного цикла проекта по модернизации;
– надежность процесса проектирования и дальнейшей эксплуатации модернизированного проекта;
– простота в ведении проектной документации, утвержденной на предприятии.

Из существующих признаков и классификаторов технологий проектирования информационных систем был выбран канонический класс технологии проектирования [Инюшкина 2014], сочетающий в себе следующие классификаторы проектных решений:

– архитектура проектируемой системы будет создаваться при помощи ручного проектирования по причине отсутствия типовых развернутых проектных решений в рамках программного комплекса организации, а также малого количества документации относительно функционирующей бизнес-логики;
– по степени использования типовых проектных решений проект относится к оригинальному проектированию, потому как архитектура работы автоматизированной отчетности создается «с нуля» в соответствии с требованиями к информационной системе;
– адаптация такого проектного решения будет происходить за счет реконструкции уже существующего на основе соответствующих программных компонентов [Карючин, Мингазов, Пестова 2022].

В рамках процесса по обновлению автоматизированной отчетности необходимо использовать вычислительный кластер – набор соединенных между собой серверов организации, функционирующих вместе в рамках единой системы. Базовая организация использует быстродействующую локальную сеть для настройки такого соединения, а также все узлы вычислительного кластера используют одинаковое оборудование и одну и ту же операционную систему [Реснянская 2021].

Определяемый вычислительный кластер оперирует функционированием следующего программного обеспечения: “ST Chicago”, «ST Мобильная Торговля»[2], «ST Шаттл», «ST Репликация»[3], хранимые на сервере организации SQL-процедуры, применяемые для репликации полученных рассчитанных данных о движении товаров по торговым точкам и складам, «ST СОД».

Для изменения настроек, используемых по умолчанию в различных частях программного продукта, необходимо конфигурирование программного обеспечения, позволяющее расширять стандартные функциональные возможности приложения, настраивать рабочую среду и создавать свои собственные отчеты.

Процесс настройки построен на основе следующего алгоритма:
– создание новой хранимой процедуры, сохраняющей результаты работы во временной таблице, добавление в базу данных;

[2] ST Мобильная Торговля – Приложение для организации выездной работы и контроля полевых сотрудников // ST Systech. URL: https://sys4tec.com/products/st-mobile/ (дата обращения 12.05.2023).

[3] ST Репликация – Компонеты программного комплекса “ST Chicago”. URL: http://www.lbs-soft.ru/products/88/493/ (дата обращения 12.05.2023).

– формирование шаблона отчета с данной процедурой;
– выбор отчета в пользовательском интерфейсе, настройка использования созданной процедуры.

В рамках конфигурирования основной части продукта “ST Chicago” была создана хранимая процедура, функционирующая по следующему алгоритму:

– определение параметров для выборки данных (дата для построения отчета, группа ассортимента, организационная позиция);
– присоединение ранее созданных справочных таблиц, выборка необходимых данных для построения отчета.

Процедура предназначена для извлечения информации о товарах и связанных с ними атрибутах из базы данных, а также для подсчета количества уникальных товаров в определенном контексте. Основная функция – это выполнение выборки данных из нескольких таблиц, таких как “refRoutes”, “docJournal”, “dhSales”, “drSales”, “refGoods”, содержащих информацию о маршрутах, продажах, торговых точках, классификации товаров. Реализацией условия для подсчета товаров является использование оператора “case”, в данном случае процедура оценивает значение переменной “@assortiment_group” (группа товаров, участвующих в акции) и, в зависимости от результата, подсчитывает количество уникальных товаров в установленном контексте или возвращает 0, если условие не соответствует. Итогом обработки являются выбранные атрибуты и подсчитанное количество уникальных товаров, которые предоставляются в итоговом наборе данных.

Итоговый вид разработанного отчета отображен на рис. 4.

В рамках рассматриваемой задачи по конфигурированию программного комплекса “ST Chicago”, а также составляющих модулей, необходимых для процесса произведения переноса данных от дистрибьютора в централизованную базу данных, был сформирован централизованный тип технологического процесса обработки информации.

Обработка информации производится исключительно на центральном сервере организации, за счет выполнения обращений и запуска хранимых процедур главного программного продукта. Репликация исходных данных производится автоматически при обнаружении необходимых файлов в папках дистрибьютора, результат по обработке информации виден как сотрудникам отдела системной аналитики, так и сотрудникам дистрибьютора.

Для решения подзадачи по настройке отображения информации в автоматизированных отчетах при помощи подключения к MS Excel была сформирована хранимая процедура, получающая данные по следующему алгоритму:

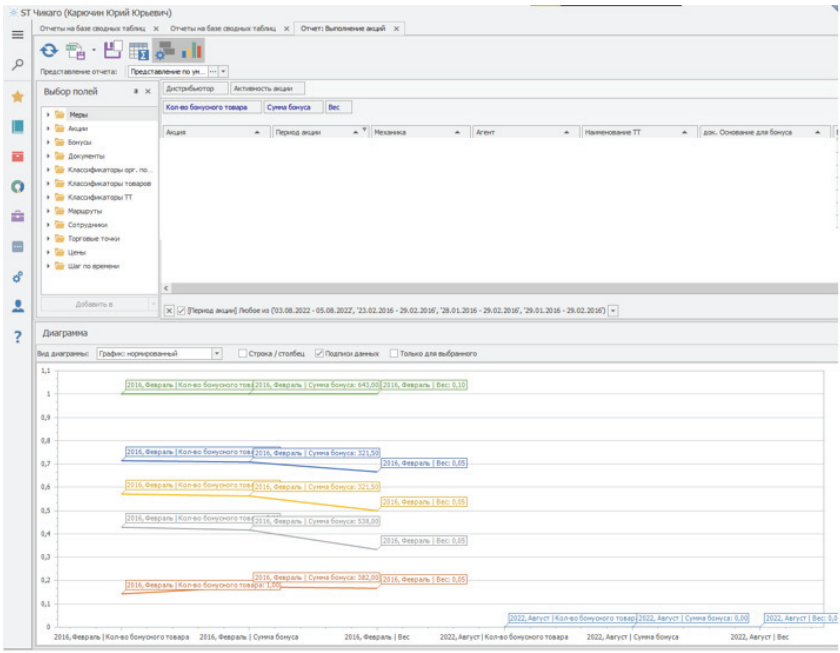


Рис. 4. Итоговый вид отчета
по всем типам акционных мероприятий

– определение даты для формирования отчета (отдельно год, отдельно месяц);
– выборка всей информации по реализации на основе определенных переменных, записывающих в себя год и месяц;
– выполнение связывания товаров, проданных по соответствующему номеру заказа и фактических продаж, переданных ранее при помощи программного комплекса “ST Chicago”;
– запись полученного массива данных в таблицу фактов.

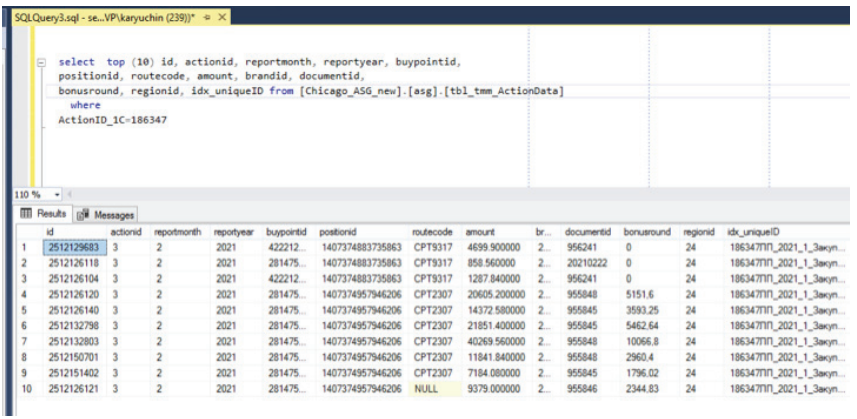
Процедура используется для извлечения массива данных о записях продаж, с применением системы бонусов и скидок для акционных мероприятий, сбора и передачи информации о каналах реализации продукции, а также состава каждой отдельной товарно-транспортной накладной, подходящей под согласованные условия трейд-маркетинговых акционных мероприятий. В данном контексте используются таблицы “#ChicagoDocsRelations”, “#SalesDoc”, “Dbo.drSales”, “dbo.refGoods” и “refattributesvalues”, хранящие в себе каналы продаж товаров, номера и составы доку-

ментов реализации продукции, а также дополнительные атрибуты, подтверждающие участие торговой точки в акции. Итогом выборки является расчетная часть описанной выше процедуры, а именно агрегация и выборка данных из таблиц “Dbo.drSales” и “dbo.refGoods” по ранее выбранным критериям внутри программного комплекса “ST Chicago” – меню создания акционных мероприятий.

По итогу запуска рассмотренной процедуры мы получаем двумерную таблицу, в которой отображены следующие обязательные поля:

- идентификатор записи из переданного отчета от торгового представителя;
- тип акционного мероприятия, в рамках которого был продан товар;
- месяц продажи данного товара;
- год продажи товара;
- идентификатор торговой точки, где была произведена реализация товара;
- номер маршрута для данной торговой точки;
- цена проданных товаров в денежном эквиваленте;
- номер товарно-транспортной накладной, по которой был продан товар;
- количество бонусных рублей для выплаты за проведение трейд-маркетингового мероприятия;
- идентификатор региона для торговой точки;
- наименование отчета, в котором были переданы строки по продаже.

Пример обращения к сформированной таблице отображен на рис. 5.



	id	actionid	reportmonth	reportyear	buypointid	positionid	routecode	amount	br...	documentid	bonusround	regionid	idx_uniqueID
1	2512129683	3	2	2021	422212...	1407374883735863	CPT9317	4699.900000	2...	956241	0	24	186347ПП_2021_1_Закуп...
2	2512126118	3	2	2021	281475...	1407374883735863	CPT9317	858.560000	2...	20210222	0	24	186347ПП_2021_1_Закуп...
3	2512126104	3	2	2021	422212...	1407374883735863	CPT9317	1287.840000	2...	956241	0	24	186347ПП_2021_1_Закуп...
4	2512126120	3	2	2021	281475...	1407374957946206	CPT2307	20605.200000	2...	955848	5151.6	24	186347ПП_2021_1_Закуп...
5	2512126140	3	2	2021	281475...	1407374957946206	CPT2307	14372.580000	2...	955845	3593.25	24	186347ПП_2021_1_Закуп...
6	2512132798	3	2	2021	281475...	1407374957946206	CPT2307	21851.400000	2...	955845	5462.64	24	186347ПП_2021_1_Закуп...
7	2512132803	3	2	2021	281475...	1407374957946206	CPT2307	40269.560000	2...	955848	10066.8	24	186347ПП_2021_1_Закуп...
8	2512150701	3	2	2021	281475...	1407374957946206	CPT2307	11841.840000	2...	955848	2960.4	24	186347ПП_2021_1_Закуп...
9	2512151402	3	2	2021	281475...	1407374957946206	CPT2307	7184.080000	2...	955845	1796.02	24	186347ПП_2021_1_Закуп...
10	2512126121	3	2	2021	281475...	1407374957946206	NULL	9379.000000	2...	955846	2344.83	24	186347ПП_2021_1_Закуп...

Рис. 5. Фрагмент экрана с результатом обращения к таблице

По итогам проведенных доработок с использованием стандартного функционала программных продуктов, развернутых в рамках предприятия, а также создания процедур передачи данных был сформирован обновленный бизнес-процесс – Архитектуры ТМА на базе “ST Chicago” («Как надо») (TO-BE).

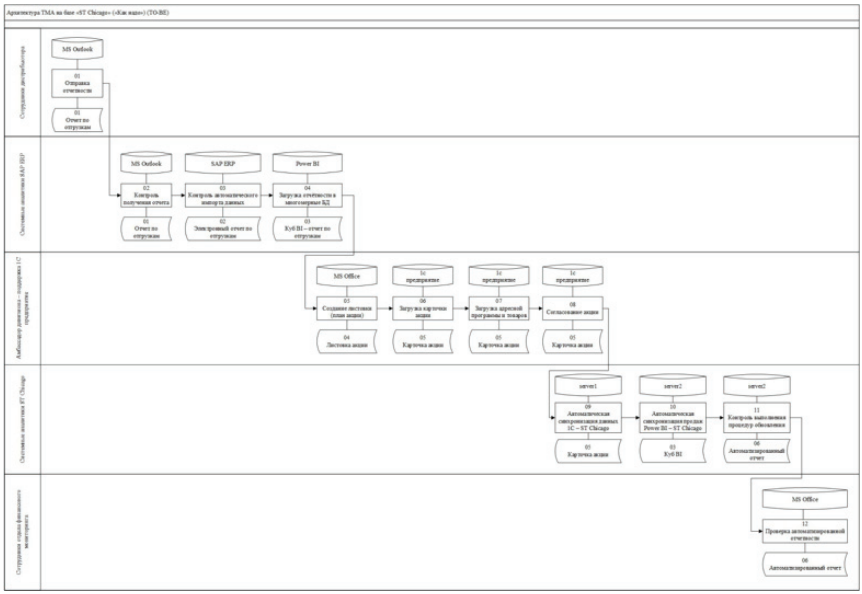


Рис. 6. «Архитектура ТМА на базе системы “ST Chicago”» («Как надо») (TO-BE)

В результате изучения и модернизации бизнес-процессов ведения трейд-маркетинговых акционных мероприятий был проведен всесторонний анализ существующей системы «Как есть» обмена данными реализации продукции, оптимизация в процесс «Как надо», с применением функционала программного комплекса “ST Chicago”. Описаны характеристики используемой в процессе обработки информации, на основе которых проведена доработка программного комплекса по управлению автоматизированной дистрибуцией с использованием существующего функционала, произведено конфигурирование используемых модулей программного продукта.

Созданные хранимые процедуры на языке структурированных запросов (SQL) позволили оптимизировать ранее автоматизи-

рованные аналитические процессы, повысить эффективность и точность управленческой отчетности с учетом указанных проектных требований. Новая модель ведения аналитики о коммерческой деятельности позволит обновлять автоматизированные отчеты в реальном времени и предоставит пользователям больший функционал для ведения трейд-маркетинговых мероприятий.

Литература

Борзов 2020 – *Борзов Д.И.* Современные тенденции в управлении маркетинговой деятельностью производственного предприятия // Вестник Российского нового университета. Серия: Человек и общество. 2020. № 1. С. 9–14.

Инюшкина 2014 – *Инюшкина О.Г.* Проектирование информационных систем (на примере методов структурного системного анализа): Учебное пособие. Екатеринбург: Форт-Диалог Исеть, 2014. 240 с.

Карючин, Мингазов, Пестова 2022 – *Карючин Ю.Ю., Мингазов Т.И., Пестова С.Ю.* Внедрение функционала и оценка эффективности программного комплекса “ST Chicago” для ведения трейд-маркетинговых мероприятий // Цифровизация и кибербезопасность: современная теория и практика: Сборник научных трудов по материалам II Международной научно-практической конференции. Омск, 2022. С. 69–73.

Козлов, Лекае, Шаповалова 2019 – *Козлов А.Д., Лекае В.А., Шаповалова М.С.* Методы анализа предметных областей: Учебное пособие. М.: РГГУ, 2019. 203 с.

Реснянская 2021 – *Реснянская Е.А.* Автоматизация и управление бизнес-процессами в компании: проблемы и решения // Вестник Российского нового университета. Серия: Человек и общество. 2021. № 3. С. 87–95.

References

Borzov, D.I. Chepurova, I.F. and Gladysheva, A.V. (2020), “Modern trends in the management of marketing activities of a production enterprise”, *Russian New University Bulletin. “Man and Society” Series,* no. 1, pp. 9–14.

Inyushkina, O.G. (2014), *Proektirovanie informatsionnykh sistem (na primere metodov strukturnogo sistemnogo analiza): Uchebnoe posobie* [Design of system information (based on methods of structural system analysis). A study guide], Fort-Dialog Iset’, Ekaterinburg, Russia.

Karyuchin, Yu.Yu., Mingazov, T.I. and Pestova, S.Yu. (2022), “The functional’s implementation and evaluation of the effectiveness of the ST Chicago software package for conducting trade marketing activities”, *Tsifrovizatsiya i kiberbezopasnost’: sovremennaya teoriya i praktika: Sbornik nauchnykh trudov po materialam*

II Mezhdunarodnoi nauchno-prakticheskoi konferentsii [Digitalization and cybersecurity. Modern theory and practice: Collection of scientific papers based on materials from the II International Scientific and Practical Conference], Omsk, Russia, pp. 69-73.

Kozlov, A.D., Lekae, V.A. and Shapovalova, M.S. (2019), *Metody analiza predmetnykh oblastei*: *Uchebnoe posobie* [Methods for analyzing subject areas. A study guide], RSUH, Moscow, Russia, 203 p.

Resnyanskaya, E.A. (2021), "Automation and management of business processes in a company. Issues and solutions", *Russian New University Bulletin, "Man and Society" Series,* no. 3, pp. 87–95.

Информация об авторах

Елена О. Шершнева, кандидат технических наук, Сибирский государственный автомобильно-дорожный университет, Омск, Россия; 644050, Россия, Омск, пр. Мира, д. 5, helen_volf@mail.ru

Юрий Ю. Карючин, магистрант, Сибирский государственный автомобильно-дорожный университет, Омск, Россия; 644050, Россия, Омск, пр. Мира, д. 5, yurykaruchin@gmail.com

Information about the authors

Elena O. Shershneva, Cand. of Sci. (Computer Science), The Siberian State Automobile and Highway University, Omsk, Russia; bld. 5, Mira lane, Omsk, 644050, Russia; helen_volf@mail.ru

Yurii Yu. Karyuchin, master student, The Siberian State Automobile and Highway University, Omsk, Russia; bld. 5, Mira lane, Omsk, 644050, Russia; yurykaruchin@gmail.com

УДК 004.8
DOI: 10.28995/2686-679X-2024-2-58-72

Система интеллектуального анализа текстов и выявления элементов информационного противоборства

Леонид Е. Алексеев
Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, alekseevle@student.bmstu.ru

Алексей Э. Самохвалов
Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, samox@bmstu.ru

Евгения Р. Смолина
Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, smolinaer@student.bmstu.ru

Аннотация. В Доктрине информационной безопасности Российской Федерации перечислены актуальные меры обеспечения информационной безопасности, в том числе информационно-аналитические технологии прогнозирования и обнаружения информационных угроз. В ходе проведения исследований методов искусственного интеллекта авторы статьи поставили перед собой цель разработать электронный онлайн-ресурс для выявления в текстовых документах признаков информационного противоборства, обнаружения ложного и вредоносного контента. В работе описан процесс подготовки обучающего набора данных и представлены результаты применения моделей машинного обучения: логистической регрессии, нейронной сети LSTM, сверточной нейронной сети Conv1D. Создано веб-приложение с API-интерфейсом для организации электронного взаимодействия с внешними информационными системами в режиме онлайн. Теоретическая значимость статьи состоит в представленной методике подготовки, обработки текстовых сообщений и визуализации итогов распознавания в них информационного противоборства. Практическая значимость работы заключается в том, что результаты выполненного исследования могут использоваться в качестве базы исследовательской, аналитической и проектной деятельности, направленной на развитие информационно-аналитических средств обеспечения информационной

безопасности. Созданный авторами сайт применим для проверки текстов на наличие в них противоборства, может использоваться в составе крупных систем мониторинга и защиты от новых информационных угроз.

Ключевые слова: информационная безопасность, информационное противоборство, машинное обучение, искусственный интеллект

Для цитирования: *Алексеев Л.Е., Самохвалов А.Э., Смолина Е.Р.* Система интеллектуального анализа текстов и выявления элементов информационного противоборства // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 58–72. DOI: 10.28995/2686-679X-2024-2-58-72

The system of intellectual analysis of texts and identification of elements of information confrontation

Leonid E. Alekseev
Bauman Moscow State Technical University, Moscow, Russia,
alekseevle@student.bmstu.ru

Alexei E. Samokhvalov
Bauman Moscow State Technical University, Moscow, Russia,
samox@bmstu.ru

Evgeniya R. Smolina
Bauman Moscow State Technical University, Moscow, Russia,
smolinaer@student.bmstu.ru

Abstract. The Information Security Doctrine of the Russian Federation lists current measures to ensure information security, including information and analytical technologies for forecasting and detecting information threats. In the course of conducting research on artificial intelligence methods, the authors of the article set themselves the goal of developing an electronic online resource for identifying signs of information warfare in text documents, detecting false and malicious content. The paper describes the process of preparing a training dataset and presents the results of using machine learning models: logistic regression, LSTM neural network, convolutional neural network Conv1D. A web application with an API interface is created for organizing electronic interaction with external information systems online. The theoretical significance of the article consists in the presented methodology of the text messages preparation, processing and visualization of the results of information confrontation recognition in those messages. The practical significance of the work lies in the fact that the

results of the research can be used as a base for research, analytical and project activities aimed at the development of information and analytical tools for ensuring information security. The website created by the authors is applicable for checking texts for the presence of confrontation in them and can be used as part of large monitoring systems and protection against new information threats.

Keywords: information security, information warfare, machine learning, artificial intelligence

For citation: Alekseev, L.E., Samokhvalov, A.E and Smolina, E.R. (2024), "The system of intellectual analysis of texts and identification of elements of information confrontation", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 2, pp. 58–72, DOI: 10.28995/2686-679X-2024-2-58-72

Введение

В современном мире информация играет ключевую роль во всех областях деятельности человека. Владение ею позволяет добиться влияния в обществе, получить экономическую выгоду, манипулировать сознанием людей, что в целом усиливает конкуренцию между государствами, общественно-политическими движениями, организациями, социальными группами и частными лицами. Эта борьба порождает информационное противоборство – «соперничество социальных систем в информационно-психологической сфере по поводу влияния на те или иные сферы социальных отношений и установления контроля над источниками стратегических ресурсов, в результате которого одни участники соперничества получают преимущества, необходимые им для дальнейшего развития, а другие их утрачивают» [Манойло 2003].

Объектом информационного противоборства могут стать любые компоненты или сегменты информационно-психологического пространства, такие как массовое и индивидуальное сознание граждан, социально-политические системы и процессы, информационная инфраструктура, информационные и психологические ресурсы.

В наши дни люди все больше обращаются к Интернету как к основному источнику информации. Согласно ежегодному отчету "Global Digital 2022", 62,5% мирового населения используют Интернет[1]. Исследование Pew Research Center показало, что 8 из

[1] Bank Individuals using internet (% of population), 2021. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2020&start=1960&view=chart (дата обращения 17.11.2023).

10 американцев узнают новости из Интернета[2]. Традиционные СМИ, представленные в газетах, журналах, на телевидении и радио, уже не могут конкурировать с таким объемом информации, который ежесекундно предоставляется в Интернете. В связи с этим информационное противоборство все чаще происходит именно в глобальной сети, а объемы информации, представленные в ней, в свою очередь, затрудняют выявление противоборства для своевременного реагирования на него. Здесь также играет роль человеческий фактор – людям бывает сложно преодолеть информационный шум, который часто используется для маскировки информационно-психологических операций. По утверждению В.Н. Ремарчука, «усиление информационной зависимости человека от растущего объема потребляемой социальной информации требует упорядочения и системной организации самой информации, совершенствования используемых информационных аналитических технологий; создания механизмов эффективной защиты человека от "вредной" информации, подменяющей истинные смыслы и ценности» [Ремарчук 2022].

В докладе Е.С. Зиновьевой «Тенденции развития информационного противоборства в условиях глобальной информатизации» на VI Всероссийской научно-практической конференции «Информационная безопасность: вчера, сегодня, завтра» (РГГУ, 12 апреля 2023 г.) отмечено, что «развитие "цифровых" технологий порождает новые риски и проблемы, связанные с обострением внешних и внутренних угроз в информационной сфере для личности, общества и государства. Противоборство в информационной сфере между государственными, общественными, коммерческими структурами, социальными и политическими группами по мере развития "цифровых" технологий обостряется до крайних форм своего проявления – информационных войн и информационного терроризма» [Арутюнов, Гришина 2023].

Авторы статьи провели исследование технологий машинного обучения (обширный раздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться), рассмотрели ML-модели, поддерживающие методы извлечения информации и интеллектуального анализа данных [Попова, Ревунков, Гапанюк 2023], сравнили их эффективность.

Цель работы – создать электронный онлайн-ресурс для выявления в текстовых документах признаков информационного

[2] More than eight-in-ten Americans get news from digital devices, 2021. URL: https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ (дата обращения 17.11.2023).

противоборства: информационно-психологических операций, «фейковых» новостей и провокаций.

Актуальность работы

26 марта 2021 г. на заседание Совета Безопасности России, в ходе которого рассматривался проект «Основ государственной политики Российской Федерации в области международной информационной безопасности», президент В.В. Путин отметил, что «новые технологические решения порождают и новые риски. Мы видим, что глобальное цифровое пространство нередко становится площадкой для жесткого информационного противоборства, для нечестной конкуренции и кибератак. Все это качественно меняет ситуацию на международной арене. Цифровую среду используют международные террористы, организованная преступность. Словом, здесь много потенциальных угроз для общей, глобальной безопасности и для отдельных стран, в том числе их суверенитета и национальных интересов»[3].

Востребованность проектируемого авторами электронного онлайн-ресурса определена в Доктрине информационной безопасности Российской Федерации в рамках деятельности по развитию системы обеспечения информационной безопасности страны, одним из модулей которой является «совершенствование информационно-аналитических и научно-технических аспектов функционирования системы обеспечения информационной безопасности»[4].

Исследование эффективности применения методов машинного обучения в распознавании информационного противоборства

Первым этапом исследования стала подготовка обучающей выборки – датасета, который представляет собой набор данных для обучения модели, обработанных следующим образом. Создан файл в формате CSV, содержащий две колонки:

---

[3] Заседание Совета Безопасности. 26 марта 2021 года. URL: http://www.kremlin.ru/events/president/news/65231 (дата обращения 17.11.2023).

[4] Указ Президента РФ от 5 декабря 2016 г. № 646 «Об утверждении Доктрины информационной безопасности Российской Федерации» // Президент России. URL: http://www.kremlin.ru/acts/bank/41460 (дата обращения 17.11.2023).

Текст (Text) – символьная строка, содержащая новость или комментарий к ней (далее будем называть их общим словом «документ»),

Наличие (Label) – числовой признак присутствия в этой новости или в комментарии к ней элементов информационного противоборства (1 – да, 0 – нет).

В файл записаны актуальные новости с признаками наличия «0» и «1» в соотношении, близком к 1:1. Символьная строка «Текст» предварительно обработана следующим образом: выполнено приведение к единому регистру символов, удалены небуквенные символы (цифры, специальные символы) и латинские буквы, удалены стоп-слова (не несущие семантической нагрузки служебные части речи, местоимения и т. п.) [Гапанюк, Попова, Рабцевич, Кобяк, Лисин 2022].

Второй этап исследования – применение алгоритма векторизации данных, который преобразовывает текст в векторы чисел, что необходимо для использования бинарной классификации. Авторы выбрали класс CountVectorizer, который входит в состав библиотеки Sklearn для машинного обучения на языке Python. Технология обработки данных следующая:

1) на вход алгоритма python-программы поступают текстовые документы;
2) с помощью CountVectorizer формируется электронный словарь всех уникальных слов (токенов) из введенных документов;
3) с помощью CountVectorizer каждый документ трансформируется в вектор, где каждый элемент этого вектора соответствует количеству вхождений соответствующего слова в документе.

Например, для предложений «Россия – священная наша держава» и «Россия – любимая наша страна» словарь будет состоять из слов: «Россия», «священная», «наша», «держава», «любимая», «страна». Затем предложения будут преобразованы в два вектора, где каждый элемент соответствует количеству вхождений соответствующего слова в этом тексте. Таким образом, первый текст «Россия – священная наша держава» будет преобразован в вектор [1, 1, 1, 1, 0]. Второй текст «Россия – любимая наша страна» будет преобразован в вектор [1, 0, 1, 0, 1].

Третий этап исследования – выбор эффективной модели для анализа текстов на наличие в них информационного противоборства. Во-первых, авторы исследовали технологию, основанную на методе логистической регрессии. Она используется для решения задач бинарной классификации с помощью методов машинного

обучения для прогнозирования вероятности того, к какому классу принадлежит объект. Логистическая регрессия реализована в модуле LogisticRegression библиотеки Sklearn. Фрагмент кода обучения модели логистической регрессии:

```
    # Обучение модели
    def train(self, dataset):
    # Работа с обучающей выборкой.
    X_train, X_test, y_train, y_test = train_test_split(dataset['text'],
dataset['label'], test_size=0.2)
    X_train_vec = self.vectorizer.fit_transform(X_train)
    X_test_vec = self.vectorizer.transform(X_test)
    # Обучение модели логистической регрессии
    self.model.fit(X_train_vec, y_train)
    # Тестирование и сохранение
    print(f»Trained {self.MODEL_NAME}:\n Saving to {self.MODEL_
BIN}...»)
    self.save_model()
    print(f»Saved to {self.MODEL_BIN}. Tesing...»)
    accuracy = self.model.score(X_test_vec, y_test)
    print(f»Testing accuracy for {self.MODEL_NAME}:», accuracy)
```

В результате кросс-валидации модели на обучающей выборке достигнут 71% точности распознавания информационного противоборства (см. рис. 1).

```
Trained logistic_regression:
 Saving to logistic_regression.sav...
Saved to logistic_regression.sav. Tesing...
Testing accuracy for logistic_regression: 0.7142857142857143
```

Рис. 1. Точность модели логистической регрессии

Во-вторых, исследована технология LSTM (Long Short-Term Memory). Она представляет собой рекуррентную нейронную сеть, которая обладает внутренней памятью и способна моделировать долгосрочные зависимости. LSTM относится к классу моделей глубокого обучения, которые применяются для работы с последовательными данными, в том числе текстами. В отличие от логистической регрессии, эта модель способна не только обрабатывать символьные последовательности переменной длины, но

и учитывать контекст при обработке каждого слова [Землевская, Червоненкис, Верховская, Юрьев 2019].

В ходе работы была использована модель LSTM и класс Tokenizer из библиотеки Keras языка Python. Создан словарь из уникальных слов, которые встречаются в текстах, и каждому слову определен уникальный целочисленный идентификатор, который может быть использован для представления слов в виде чисел. Фрагмент кода обучения модели LSTM:

```
    def train(self, dataset: DataFrame):
    # Работа с датасетом
    X_train, X_test, y_train, y_test = train_test_split(dataset['text'],
dataset['label'], test_size=0.2)
    self.tokenizer.fit_on_texts(X_train)
    X_train = self.tokenizer.texts_to_sequences(X_train)
    X_test = self.tokenizer.texts_to_sequences(X_test)
    maxlen = 100
    X_train = pad_sequences(X_train, padding='post', maxlen=maxlen)
    X_test = pad_sequences(X_test, padding='post', maxlen=maxlen)
    # Обучение
    self.model = Sequential()
    self.model.add(Embedding(input_dim=10000, output_dim=64, in-
put_length=maxlen))
    self.model.add(LSTM(64, dropout=0.2, recurrent_dropout=0.2))
    self.model.add(Dense(1, activation='sigmoid'))
    self.model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
    self.model.fit(X_train, y_train, validation_data=(X_test, y_test),
epochs=100, batch_size=32)
    # Сохранение и тестирование
    print(f»Trained {self.MODEL_NAME}:\n Saving to {self.MODEL_
BIN}...»)
    self.save_model()
    print(f»Saved to {self.MODEL_BIN}. Tesing...»)
    _, accuracy = self.model.evaluate(X_test, y_test)
    print(f»Testing accuracy for {self.MODEL_NAME}:», accuracy)
```

После обучения LSTM модели авторы статьи получили 83% точности определения информационного противоборства, что значительно превосходит результаты обучения модели на основе логистической регрессии (см. рис. 2).

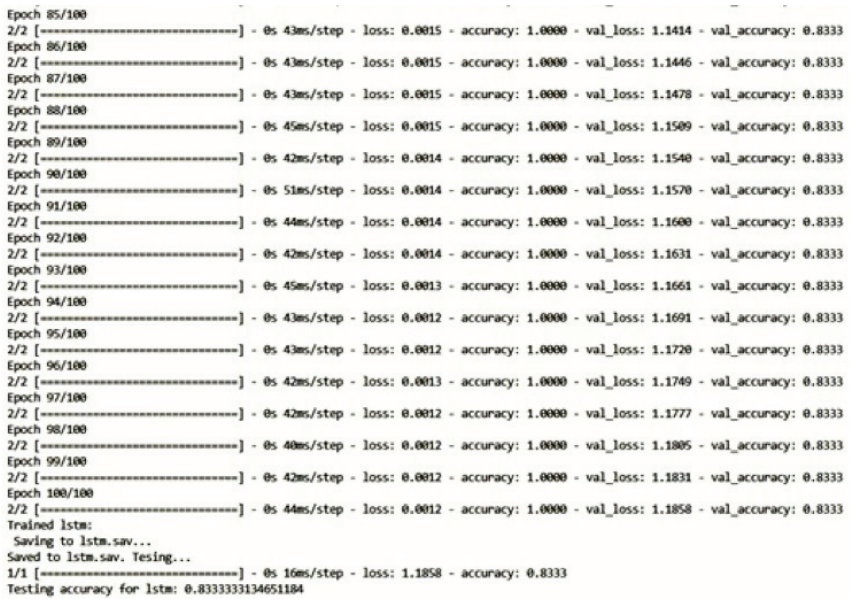


Рис. 2. Точность модели LSTM

В-третьих, исходя из полученных ранее показателей, было решено исследовать алгоритмы, которые включают в себя обучение нейронных сетей, коррекцию весов, учет последовательности слов и контекста. Conv1D представляет собой сверточную нейронную сеть, часто используемую для анализа временных рядов и текстов. Она принимает на вход последовательность данных и обрабатывает ее с помощью одномерных сверток, позволяя извлечь локальные признаки из данных. После этого входные данные проходят через пулинг-слои, уменьшающие размерность выходных данных, и обрабатываются последующими сверточными слоями [Абрамов 2023].

Фрагмент кода обучения модели Conv1D:

```
    def train(self, dataset: DataFrame):
    # Работа с датасетом
    X_train, X_test, y_train, y_test = train_test_split(dataset['text'],
dataset['label'], test_size=0.2)
    self.tokenizer.fit_on_texts(X_train)
```

```
        X_train = self.tokenizer.texts_to_sequences(X_train)
        X_test = self.tokenizer.texts_to_sequences(X_test)
        maxlen = 100
        X_train = pad_sequences(X_train, padding='post', maxlen=maxlen)
        X_test = pad_sequences(X_test, padding='post', maxlen=maxlen)
        # Инициализация, обучение
        self.model = Sequential()
        self.model.add(Embedding(input_dim=5000, output_dim=64, in-
    put_length=maxlen))
        self.model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
        self.model.add(MaxPooling1D(pool_size=2))
        self.model.add(Flatten())
        self.model.add(Dense(64, activation='relu'))
        self.model.add(Dropout(0.5))
        self.model.add(Dense(1, activation='sigmoid'))
        self.model.compile(loss='binary_crossentropy', optimizer='adam',
    metrics=['accuracy'])
        self.model.fit(X_train, y_train, validation_data=(X_test, y_test),
    epochs=100, batch_size=32)
        # Тестирование и сохранение
        print(f»Trained {self.MODEL_NAME}:\n Saving to {self.MODEL_
    BIN}...»)
        self.save_model()
        print(f»Saved to {self.MODEL_BIN}. Tesing...»)
        y_pred = self.model.predict(X_test)
        accuracy = accuracy_score(y_test, np.argmax(y_pred, axis=1))
        print(f»Testing accuracy for {self.MODEL_NAME}:», accuracy)
```

Авторы отметили такое преимущество модели Conv1D над LSTM – она содержит меньше параметров, что позволяет быстрее обрабатывать большие наборы данных. Точность распознавания с помощью Conv1D составила 85% (см. рис. 3).

Разработка веб-приложения для распознавания информационного противоборства

Авторы создали сайт http://news.kamaimedia.ru, на котором пользователи могут проверять тексты на наличие в них информационного противоборства с помощью рассмотренных моделей машинного обучения (см. рис. 4). В целях повышения точности распознавания реализован сбор пользовательских оценок, кото-

рые записываются в обучающую выборку. Для разработчиков информационных систем составлена проектная документация с описанием организации электронного взаимодействия с ресурсом с помощью HTTP-запросов.

```
Epoch 82/100
2/2 [==============================] - 0s 18ms/step - loss: 7.5300e-04 - accuracy: 1.0000 - val_loss: 0.3160 - val_accuracy: 0.8571
Epoch 83/100
2/2 [==============================] - 0s 20ms/step - loss: 3.8817e-04 - accuracy: 1.0000 - val_loss: 0.3151 - val_accuracy: 0.8571
Epoch 84/100
2/2 [==============================] - 0s 22ms/step - loss: 0.0024 - accuracy: 1.0000 - val_loss: 0.3203 - val_accuracy: 0.8571
Epoch 85/100
2/2 [==============================] - 0s 19ms/step - loss: 0.0035 - accuracy: 1.0000 - val_loss: 0.3311 - val_accuracy: 0.8571
Epoch 86/100
2/2 [==============================] - 0s 18ms/step - loss: 4.4167e-04 - accuracy: 1.0000 - val_loss: 0.3459 - val_accuracy: 0.8571
Epoch 87/100
2/2 [==============================] - 0s 19ms/step - loss: 5.2448e-04 - accuracy: 1.0000 - val_loss: 0.3600 - val_accuracy: 0.8571
Epoch 88/100
2/2 [==============================] - 0s 19ms/step - loss: 3.3839e-04 - accuracy: 1.0000 - val_loss: 0.3721 - val_accuracy: 0.8571
Epoch 89/100
2/2 [==============================] - 0s 19ms/step - loss: 7.3134e-04 - accuracy: 1.0000 - val_loss: 0.3819 - val_accuracy: 0.8571
Epoch 90/100
2/2 [==============================] - 0s 20ms/step - loss: 3.0357e-04 - accuracy: 1.0000 - val_loss: 0.3897 - val_accuracy: 0.8571
Epoch 91/100
2/2 [==============================] - 0s 18ms/step - loss: 6.4094e-04 - accuracy: 1.0000 - val_loss: 0.3951 - val_accuracy: 0.8571
Epoch 92/100
2/2 [==============================] - 0s 20ms/step - loss: 2.3293e-04 - accuracy: 1.0000 - val_loss: 0.3989 - val_accuracy: 0.8571
Epoch 93/100
2/2 [==============================] - 0s 19ms/step - loss: 3.2850e-04 - accuracy: 1.0000 - val_loss: 0.4010 - val_accuracy: 0.8571
Epoch 94/100
2/2 [==============================] - 0s 19ms/step - loss: 2.0571e-04 - accuracy: 1.0000 - val_loss: 0.4023 - val_accuracy: 0.8571
Epoch 95/100
2/2 [==============================] - 0s 18ms/step - loss: 3.0579e-04 - accuracy: 1.0000 - val_loss: 0.4031 - val_accuracy: 0.8571
Epoch 96/100
2/2 [==============================] - 0s 19ms/step - loss: 6.9406e-04 - accuracy: 1.0000 - val_loss: 0.4028 - val_accuracy: 0.8571
Epoch 97/100
2/2 [==============================] - 0s 18ms/step - loss: 3.3781e-04 - accuracy: 1.0000 - val_loss: 0.4015 - val_accuracy: 0.8571
Epoch 98/100
2/2 [==============================] - 0s 19ms/step - loss: 5.5380e-04 - accuracy: 1.0000 - val_loss: 0.3991 - val_accuracy: 0.8571
Epoch 99/100
2/2 [==============================] - 0s 18ms/step - loss: 3.0591e-04 - accuracy: 1.0000 - val_loss: 0.3966 - val_accuracy: 0.8571
Epoch 100/100
2/2 [==============================] - 0s 19ms/step - loss: 3.4266e-04 - accuracy: 1.0000 - val_loss: 0.3931 - val_accuracy: 0.8571
```

Рис. 3. Точность модели Conv1D

Рис. 4. Интерфейс веб-приложения

Авторы спроектировали представленное приложение как открытую систему. С помощью интерфейса API (Application Programming Interface, см. рис. 5) разработчики информационных систем, аналитики и иные заинтересованные лица могут анализировать текстовые документы с применением реализованных моделей машинного обучения. Благодаря поступающим по этому каналу текстам и их пользовательским (экспертным) оценкам выборка для обучения постоянно увеличивается, что позволяет более точно определять информационное противоборство. Таким образом можно создать систему мониторинга публичных Интернет-ресурсов, способную анализировать тексты в автоматическом онлайн-режиме.

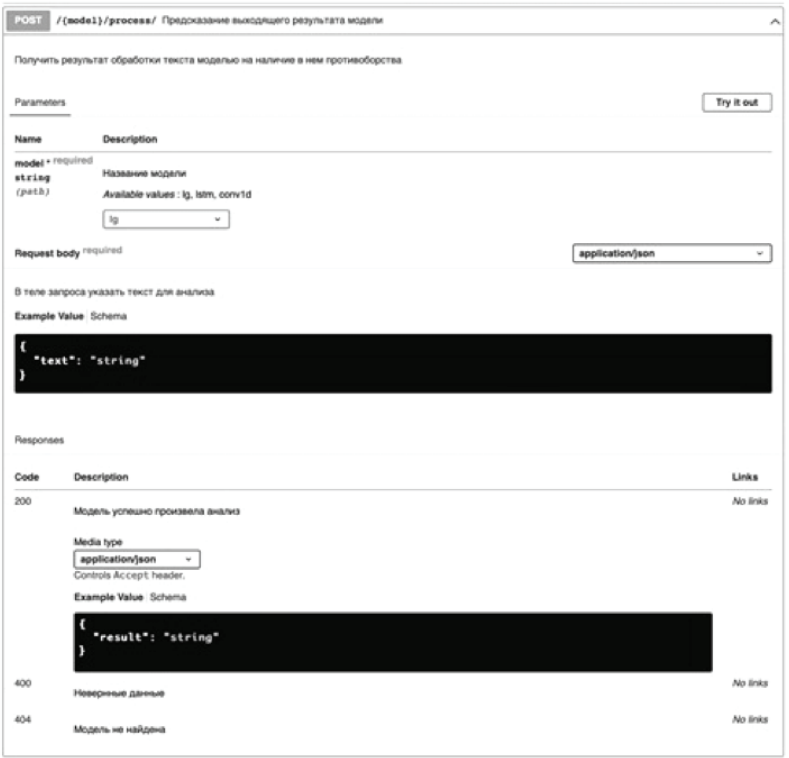


Рис. 5. Application Programming Interface веб-приложения

Заключение

Технологии искусственного интеллекта позволяют выполнять поиск элементов информационного противоборства в электронных документах, публикациях, сообщениях, блогах в сети Интернет с высокой точностью. При применении нейронных сетей LSTM и Conv1D вероятность распознавания ложного и вредоносного контента превышает 85%.

Разработанное веб-приложение представляет собой программную реализацию моделей машинного обучения на языке Python. Авторы уверены, что организация обратной связи от пользователей и хранение их экспертных оценок вводимых в систему текстов обеспечит регулярное пополнение обучающей выборки и повысит эффективность методов распознавания.

Бурное внедрение искусственного интеллекта привело к появлению нового поколения систем – гибридных интеллектуальных информационных систем (ГИИС), в которые «встраиваются модули интеллектуальной обработки данных и знаний» [Черненький, Терехов, Гапанюк 2016]. Представленная в этой статье система обнаружения информационного противоборства может быть отнесена к классу таких современных систем.

Литература

Абрамов 2023 – *Абрамов Д.А.* Применение сверточных сетей в задаче классификации текста // Научный аспект. 2023. Т. 6. № 2. С. 587–594.

Арутюнов, Гришина 2023 – *Арутюнов В.В., Гришина Н.В.* Об итогах VI Всероссийской научно-практической конференции «Информационная безопасность: вчера, сегодня, завтра» // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 3. С. 38–48. DOI: 10.28995/2686-679X-2023-3-38-48.

Гапанюк, Попова, Рабцевич, Кобяк, Лисин 2022 – *Гапанюк Ю.Е., Попова Д.А., Рабцевич К.Р., Кобяк А.В., Лисин А.В.* Классификация текстов различными методами машинного обучения для выявления признаков киберагрессии // Естественные и технические науки. 2022. № 6 (169). С. 277–281.

Землевская, Червоненкис, Верховская, Юрьев 2019 – *Землевская А.С., Червоненкис М.А., Верховская Е.К., Юрьев Г.А.* Разработка системы автоматической классификация текстов при помощи LSTM-модели // Нейрокомпьютеры и их применение: XVII Всероссийская научная конференция. Тезисы докладов, Москва, 19 марта 2019 г. М.: Московский государственный психолого-педагогический университет, 2019. С. 301–302.

Манойло 2018 – *Манойло А.В.* Государственная информационная политика в особых условиях. М.: МИФИ, 2003. 388 с.

Попова, Ревунков, Гапанюк 2023 – *Попова И.А., Ревунков Г.И., Гапанюк Ю.Е.* AutoML: исследование существующих программных реализаций и определение общей внутренней структуры решений // Труды Института системного анализа Российской академии наук. 2023. Т. 73. № 1. С. 43–54. DOI: 10.14357/20790279230106.

Ремарчук 2022 – *Ремарчук В.Н.* Информационная аналитика: теория, методология, технологии: Учебник для вузов. СПб.: Лань, 2022. 224 с.

Черненький, Терехов, Гапанюк 2016 – *Черненький В.М., Терехов В.И., Гапанюк Ю.Е.* Структура гибридной интеллектуальной информационной системы на основе метаграфов // Нейрокомпьютеры: разработка, применение. 2016. № 9. С. 3–14.

References

Abramov, D.A. (2023), "Application of convolutional networks in the problem of text classification", *Scientific aspect*, vol. 6, no. 2. pp. 587–594.

Arutyunov, V.V. and Grishina, N.V. (2023), "On the results of the 6th All-Russian Scientific and Practical Conference 'Information Security. Yesterday, Today, Tomorrow' ", *RSUH/RGGU Bulletin. "Information Scince. Information Security. Mathematics" Series*, no. 3, pp. 38–48, DOI: 10.28995/2686-679X-2023-3-38-48.

Chernenky, V.M., Terekhov, V.I. and Gapanyuk, Yu.E. (2016), "The structure of a hybrid intelligent information system based on metagraphs", *Neurocomputers: development, application*, no. 9, pp. 3–14.

Gapanyuk, Yu.E., Popova, D.A., Rabtsevich, K.R., Kobyak, A.V. and Lisin, A.V. (2022), "Classification of texts by various machine learning methods to identify signs of cyber aggression", *Natural and technical sciences*, no. 6 (169), pp. 277–281.

Manoilo, A.V. (2018), *Gosudarstvennaya informatsionnaya politica v osobykh usloviyakh* [State information policy in special conditions], MEPHI, Moscow, Russia.

Popova, I.A., Revunkov, G.I. and Gapanyuk, Yu.E. (2023), "AutoML. Investigation of existing software implementations and determination of the general internal structure of solutions", *Proceedings of the Institute of System Analysis of the Russian Academy of Sciences*, vol. 73, no. 1, pp. 43–54, DOI: 10.14357/20790279230106.

Remarchuk, V.N. (2022), "Informatsionnaya analitika: teoriya, metodologiya, tekhnologii" [Information analytics. Theory, methodology, technology], Lan', St. Petersburg, Russia.

Zemlevskaya, A.S., Chervonenkis, M.A., Verkhovskaya, E.K. and Yur'yev, G.A. (2019), "Development of an automatic text classification system using the LSTM model", *Neirokomp'yutery i ikh primenenie: XVII Vserossiiskaya nauchnaya konferentsiya. Tezisy dokladov, Moskva, 19 marta 2019 g.* [Neurocomputers and their application. The 17th All-Russian Scientific Conference. Abstracts of reports, Moscow, March 19, 2019], Moscow State Psychological and Pedagogical University, Moscow, Russia, pp. 301–302.

Информация об авторах

Леонид Е. Алексеев, студент, Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; alekseevle@student.bmstu.ru

Алексей Э. Самохвалов, кандидат экономических наук, Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; samox@bmstu.ru

Евгения Р. Смолина, студент, Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; smolinaer@student.bmstu.ru

Information about the authors

Leonid E. Alekseev, student, Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, 105005, Russia; alekseevle@student.bmstu.ru

Alexei E. Samokhvalov, Cand. of Sci. (Economics), Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, 105005, Russia; samox@bmstu.ru

Evgenia R. Smolina, student, Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, 105005, Russia; smolinaer@student.bmstu.ru

УДК 004.056
DOI: 10.28995/2686-679X-2024-2-73-82

Анализ подходов к расследованию инцидентов информационной безопасности

Наталия В. Гришина
Российский государственный гуманитарный университет, Москва, Россия;
Московский государственный лингвистический университет, Москва, Россия, grnat@rambler.ru

Аннотация. В статье проводится аналитический обзор материалов по расследованию инцидентов информационной безопасности. Рассмотрено определение понятия инцидента информационной безопасности. Приведены нормативные документы, регламентирующие деятельность, связанную с инцидентами информационной безопасности. Определены основные направления и подходы к рассмотрению инцидентов.

Автор обращает внимание на тот факт, что такие понятия, как «информационная безопасность» и «кибербезопасность», зачастую используют как синонимы. Понятие «информационная безопасность» представляется более широким по отношению к понятию «кибербезопасность». Кибербезопасность может быть определена как одна из составляющих обеспечения информационной безопасности и направлена на защиту от атак в киберпространстве. Следовательно и проблема расследования инцидентов информационной безопасности должна быть рассмотрена именно в широком понимании, поскольку инциденты информационной безопасности могут быть реализованы не только в рамках киберпространства.

Вся деятельность по сбору информации об инцидентах, предупреждению инцидентов, локализации последствий и расследованию инцидентов должна быть направлена на объект информатизации в целом. В статье намечены основные направления дальнейшего исследования данной проблемы. Особое внимание уделяется проблеме «человеческого фактора» в связи.

Ключевые слова: информационная безопасность, защита информации, система защиты информации, кибербезопасность, инцидент информационной безопасности, человеческий фактор

Для цитирования: Гришина Н.В. Анализ подходов к расследованию инцидентов информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 73–82. DOI: 10.28995/2686-679X-2024-2-73-82

Analysis of approaches to investigating information security incidents

Natalia V. Grishina
Russian State University for the Humanities, Moscow, Russia;
Moscow State Linguistic University, Moscow, Russia,
grnat@rambler.ru

Abstract. The article presents an analytical review of materials on the investigation of information security incidents. It considers the definition of the concept of an information security incident. Regulatory documents regulating activities related to information security incidents are provided and the main directions and approaches to handling incidents are identified.

The author draws attention to the fact that concepts such as information security and cybersecurity are often used as synonyms. The concept of information security seems broader in relation to the concept of cybersecurity. Cybersecurity can be defined as one of the components of information security and is aimed at protecting against attacks in cyberspace. Consequently, the issue of investigating information security incidents should be considered in a broad sense, since information security incidents can occur not only within cyberspace.

All activities to collect information about incidents, their prevention, containment and investigation should be aimed at the informatization object as a whole. The article outlines the main directions for further research of the issue. Particular attention is paid to the "human factor" issue in communication.

Keywords: information security, information protection, information security system, cybersecurity, information security incident, human factor

For citation: Grishina, N.V. (2024), "Analysis of approaches to investigating information security incidents", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 2, pp. 73–82, DOI: 10.28995/2686-679X-2024-2-73-82

В общей структуре проблемы управления информационной безопасностью (ИБ) значимое место занимают вопросы проявления и расследования инцидентов нарушения информационной безопасности.

В международном стандарте ISO/IEC 27000:2014 инцидент определяется как «событие или серия нежелательных или непредвиденных событий ИБ, которые могут с большой долей вероятности привести к компрометации бизнес-операций или созданию

угрозы ИБ»[1]. В ФЗ-149 сказано, что необходимо «своевременное обнаружение фактов несанкционированного доступа к информации». Здесь речь идет, по сути, также об управлении инцидентами.

Инцидент информационной безопасности можно определить как отклонение от нормы, заданной в политике корпоративной безопасности. В общем случае возникновение инцидентов можно связать с отклонениями в функционировании системы защиты информации или с ее несовершенством. Если подобный инцидент произошел, необходимо проанализировать причины, источники и последствия его возникновения и принять меры по минимизации вероятности возникновения подобных инцидентов в будущем.

Анализ инцидентов информационной безопасности можно рассматривать с различных точек зрения:

- правового обоснования;
- принципов и подходов к расследованию соответствующих инцидентов;
- основных этапов расследования;
- выявление виновных;
- ликвидация последствий реализованного инцидента.

Цель статьи – провести анализ известных подходов к расследованию инцидентов информационной безопасности на объекте информатизации и сформировать обобщенный взгляд на указанную проблему.

В статье Д.Н. Шевченко [Шевченко 2020] рассмотрены основные понятия, связанные с цифровой криминалистикой. Показано, что методы компьютерной криминалистики обеспечивают правовую базу для расследования инцидентов информационной безопасности.

В статье Н.Г. Лабутина и П.В. Костина [Лабутин, Костин 2021] раскрываются основные понятия, связанные с инцидентами информационной безопасности. Предложены подходы и методы обнаружения данных о действиях нарушителей. Рассмотрены правила поведения сотрудников, направленные на сохранение «следов» действия нарушителей. Раскрыты типичные ошибки, которые могут проявляться при расследовании инцидентов ИБ. Особое внимание уделяется вопросам своевременного обнаружения инцидентов в целях оперативного реагирования.

---

[1] ISO/IEC 27000:2014 (п. 2.36): Information security incident – single or a series of unwanted or unexpected information security events that have a significant probability of compromising business operations and threatening information security. URL: https://cdn.standards.iteh.ai/samples/63411/888e9d5783ca4a63ae66092fe57c2e7c/ISO-IEC-27000-2014.pdf (дата обращения 12.12.2023).

Многие авторы обсуждают возможность использования регистрационных журналов для сбора и последующей обработки информации [Курдюков, Твердова 2023]. В статье показаны типы систем, в функции которых входит накопление данных о событиях в системе и описывается порядок действий по сбору и анализу событий.

Тема использования регистрационной информации не нова. На целесообразность постобработки регистрационной информации указывалось в работах [Гришина 2021; Рытов, Калашников 2019]. Методы постобработки можно использовать как стратегический ресурс [Гришина, Емельянов 2006].

Большая часть исследований, представленных в научной литературе, относится к расследованию киберпреступлений. Научные исследования на тему расследования киберпреступлений можно рассмотреть на примере целого ряда статей.

К теме разработки новых методик и тактик для расследования различных видов киберпреступлений обращается А.Б. Смушкин [Смушкин 2023]. Автор рассматривает использование искусственного интеллекта в качестве инструмента для расследования киберпреступлений, однако выделяет тот факт, что основное место в расследовании занимает именно человек, который должен пользоваться определенной методикой. Причем методика должна соответствовать не только специфике киберпреступлений в целом, но и преступлений, совершенных с использованием конкретных информационно-технологических устройств или концептов. Целесообразно использовать методы автоматизации в процессе расследования инцидентов [Яковлева 2023].

Некоторые авторы [Левшун, Гайфулина, Чечулин, Котенко 2020] предлагают проводить классификацию злоумышленников, реализовавших инцидент ИБ по различным классификационным признакам с целью определения их целей и действий. В то же время конкретные методы определения атаки в статье отсутствуют.

Авторы статьи [Alastal, Shaqfa 2023; Косенкова, Романовский, Цацкина 2023] обращают внимание на нанесенный ущерб различным сферам жизни человека киберпреступностью. Также авторы раскрывают основные проблемы, с которыми сталкиваются полицейские, расследующие киберпреступления, среди которых выделяют:

– несоответствие между темпами развития киберпреступности и скоростью расследования данного вида преступлений;
– недостаточные подготовка и осведомленность об этих быстро развивающихся видах преступлений;
– неспособность идти в ногу с технологическими изменениями.

Исходя из этого следует вывод – без использования современных и эффективных методик расследования киберпреступлений невозможна эффективная работа полиции в сфере их расследования.

В своей статье [Hamad, Derar 2022] авторы, проведя обзор и систематизацию программных средств, предназначенных для расследования киберпреступлений по различным категориям и видам, приходят к выводу о том, что полностью интегрированного инструмента для расследования не существует.

В статье [Никитин 2022] автор подчеркивает, что если направить усилия на предупреждение инцидентов информационной безопасности, то отпадет и необходимость расследования.

Целый ряд статей [Рытов, Калашников 2019; Наврузов 2022] содержат исследования по идентификации инцидентов и методы сбора и анализа данных по этим инцидентам.

В результате проведенного анализа предметной области можно сделать следующие выводы:

- авторы статей анализируют частные случаи и не делают обобщений относительно инцидентов ИБ;
- отсутствуют универсальные подходы к обоснованию превентивных мер противодействия киберпреступлениям;
- при расследовании преступлений исследователи предлагают использовать профайлинг личности [Акименко, Овчаренко 2022; Русецкая 2022];
- авторы подчеркивают, что эффективность расследования киберпреступлений в значительной мере зависит от знаний и опыта специалиста по информационной безопасности;
- практически все авторы сходятся во мнении, что в ближайшей перспективе следует создавать комплексные системы по противодействию киберпреступлениям;
- необходимо планировать ресурсы, выделяемые на расследование инцидентов информационной безопасности.

Особое внимание следует обратить на тот факт, что такие понятия, как «информационная безопасность» и «кибербезопасность» зачастую используют как синонимы. Надо отметить, что это не совсем верно. Понятие «информационная безопасность» представляется нам более широким по отношению к понятию «кибербезопасность». Кибербезопасность может быть определена как одна из составляющих информационной безопасности, которая направлена на защиту от атак в киберпространстве.

Следовательно, проблема расследования инцидентов информационной безопасности должна быть рассмотрена именно в широком понимании. Инциденты информационной безопасности могут быть реализованы не только в рамках киберпространства.

Человек является неотъемлемым звеном объекта информатизации. С точки зрения реализации инцидента ИБ роль человека переоценить невозможно. По различным оценкам, именно с деятельностью работника организации связано до 80% всех инцидентов ИБ объекта информатизации.

Заключение

1. Анализ и расследование инцидентов необходимо рассматривать относительно объекта информатизации.
2. Анализируя частные случаи, надо делать обобщения относительно инцидентов информационной безопасности с целью формирования базы инцидентов и методов противодействия им.
3. Организация системы нейтрализации факторов, вызывающих инциденты, даст возможность обеспечить защиту от широкого класса угроз.
4. Использование методов искусственного интеллекта позволит повысить результативность системы управления инцидентами информационной безопасности.
5. Тщательный профайлинг личности целесообразно проводить не при расследовании инцидентов, а ранее, на этапе подбора специалистов; при этом необходимо обращать внимание не только на профессиональную составляющую, но и морально-психологические характеристики личности.
6. Перспективная система противодействия инцидентам ИБ должна быть комплексной; ее целесообразно формировать как составную часть системы защиты информации объекта информатизации и оснащать всеми необходимыми ресурсами.
7. В рамках совершенствования системы корпоративной безопасности следует особое внимание уделять работе с персоналом и вопросам переподготовки кадров.

Литература

Акименко, Овчаренко 2022 – *Акименко М.А., Овчаренко И.А.* К проблеме применения криминалистического профайлинга при расследовании и раскрытии киберпреступлений // Евразийская адвокатура. 2022. № 1 (56). С. 82.

Гришина, Емельянов 2006 – *Гришина Н.В., Емельянов С.А.* Деловая разведка как разновидность информационной работы // Прикладная информатика. 2006. № 3. С. 34–41.

Гришина, Мецатунян, Русецкая 2012 – *Гришина Н.В., Мецатунян М.В., Русецкая И.А.* Влияние социально-психологических аспектов на обеспечение информационной безопасности субъектов информационных отношений // Безопасность информационных технологий. 2012. № 1. С. 12–17.

Гришина 2021 – *Гришина Н.В.* Основы информационной безопасности предприятия: учебное пособие. М.: ИНФРА-М, 2021. 216 с.

Косенкова, Романовский, Цацкина 2023 – *Косенкова Ю.Ю., Романовский С.В., Цацкина Е.П.* Моделирование процесса оценки угроз безопасности информационных систем налоговых органов // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 2. С. 46–69.

Курдюков, Твердова 2023 – *Курдюков П.Р., Твердова С.М.* Сбор и анализ регистрационных журналов при проведении расследований инцидентов информационной безопасности // Электронный журнал: наука, техника и образование. СВ1/2023 (ИБ) (41). С. 27–30.

Лабутин, Костин 2021 – *Лабутин Н.Г., Костин П.В.* Расследование инцидентов информационной безопасности в территориальном налоговом органе ФНС России // Актуальные вопросы налогового администрирования в контексте современных тенденций профессионального развития государственных гражданских служащих: Материалы научно-практической конференции (Нижний Новгород, 26 мая 2021 г.). Н. Новгород, 2021. С. 157–163.

Левшун, Гайфулина, Чечулин, Котенко 2020 – *Левшун Д.С., Гайфулина Д.А., Чечулин А.А., Котенко И.В.* Проблемные вопросы информационной безопасности киберфизических систем // Тр. СПИИРАН. 2020. Вып. 19. Т. 5. С. 1050–1088.

Наврузов 2022 – *Наврузов Э.Р.* О формировании баз прецедентов для решения задач информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2022. № 3. С. 66–84.

Никитин 2022 – *Никитин А.М.* Разработка информационно-аналитической системы поддержки принятия решений при расследовании инцидентов информационной безопасности // Кибербезопасность: технические и правовые аспекты защиты информации: Материалы межвузовской студенческой научно-практической конференции. Волгоград, 2022. С. 189–193.

Русецкая 2022 – *Русецкая И.А.* Роль профайлинга в обеспечении информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2022. № 3. С. 85–95. DOI: 10.28995/2686-679X-2022-3-85-95.

Рытов, Калашников 2019 – *Рытов М.Ю., Калашников Р.Ю.* Система анализа данных для расследования инцидентов информационной безопасности в сети предприятия // Информационные технологии и системы: управление, экономика, транспорт, право. 2019. № 2 (34). С. 208–211.

Смушкин 2023 – *Смушкин А.Б.* Цифровая трансформация криминалистики. Ч. 2: Цифровая трансформация организации расследования преступлений, крими-

налистической тактики и методики расследования отдельных видов преступлений // Вестник СГЮА. 2023. № 2 (151). С. 272–278.

Шевченко 2020 – *Шевченко Д.Н.* Методы цифровой криминалистики и компьютерной форензики для расследования инцидентов информационной безопасности – обзор // Проблемы науки. Компьютерные и информационные науки. 2020. № 4 (52). С. 32–33.

Яковлева 2023 – *Яковлева А.О.* О вопросах применения инструментов автоматизации процесса расследования инцидентов // Актуальные проблемы авиации и космонавтики. 2023. Т. 2. С. 406–407.

Alastal, Shaqfa 2023 – *Alastal A., Shaqfa A.* Enhancing Police Officers' Cybercrime Investigation Skills Using a Checklist Tool // Journal of Data Analysis and Information Processing. 2023. Vol. 11. P. 121–143. DOI: 10.4236/jdaip.2023.112008.

Hamad, Derar 2022 – *Hamad N., Derar E.* Digital Forensics Tools Used in Cybercrime Investigation, Comparative Analysis // Journal of Xi'an University of Architecture & Technology, 2021. Vol. 23, issue 3. P. 367–379. DOI: 10.37896/JXAT14.04/314909.

References

Akimenko, M.A. and Ovcharenko, I.A. (2022), "On the issue of using forensic profiling in the investigation and detection of cybercrimes", *Eurasian Advocacy*, no. 1 (56), p. 82.

Alastal, A. and Shaqfa, A. (2023), "Enhancing Police Officers' Cybercrime Investigation Skills Using a Checklist Tool", *Journal of Data Analysis and Information Processing*, no. 11, pp. 121–143.

Grishina, N.V. (2021), *Osnovy informatsionnoi bezopasnosti predpriyatiya: Uchebnoe posobie* [Fundamentals of enterprise information security. Study guide], INFRA-M, Moscow, Russia, 216 p.

Grishina, N.V. and Emel'yanov, S.A. (2006), "Business intelligence as a type of information work", *Applied Informatics*, no. 3, pp. 34–41.

Grishina, N.V., Metsatunyan, M.V. and Rusetskaya, I.A. (2012), "The influence of sociopsychological aspects on ensuring information security of subjects of information relations", *Security of information technologies*, no. 1, pp. 12–17.

Kosenkova, Yu.Yu., Romanovskii, S.V. and Tsatskina, E.P. (2023), "Information security threats assessment process modeling for tax authorities' information systems", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 2, pp. 46–69, DOI: 10.28995/2686-679X-2023-2-46-69.

Kurdyukov, P.R. and Tverdova S.M. (2023), "Collection and analysis of logs during investigations of information security incidents", *Electronic journal: science, technology and education,* SV1/2023 (IB) (41), pp.27-30.

Labutin, N.G. and Kostin, P.V. (2021), "Investigation of information security incidents in the territorial tax authority of the Federal Tax Service of Russia", *Aktual'nye voprosy nalogovogo administrirovaniya v kontekste sovremennykh tendentsii professional'nogo razvitiya gosudarstvennykh grazhdanskikh sluzhashchikh. Materialy nauchno-prakticheskoi konferentsii (Nizhnii Novgorod, 26 maya 2021 g.)* [Current issues of tax administration in the context of modern trends in the professional development of civil servants. Proceedings of the scientific and practical conference (Nizhny Novgorod, May 26, 2021)], N. Novgorod, Russia, pp. 157–163.

Levshun, D.S., Gaifulina, D.A., Chechulin, A.A. and Kotenko, I.V. (2020), "Problematic issues of information security of cyberphysical systems", *Tr. SPIIRAN,* issue 19, vol. 5. pp. 1050–1088.

Navruzov, E.R. (2022), "On forming the precedent bases for solving problems of the information security", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 3, pp. 66–84, DOI: 10.28995/2686-679X-2022-3-66-84.

Nikitin, A.M. (2022), "Development of an information and analytical system to support decision-making in information security incidents investigation", *Kiberbezopasnost': tekhnicheskie i pravovye aspekty zashchity informatsii. Materialy mezhvuzovskoi studencheskoi nauchno-prakticheskoi konferentsii* [Cybersecurity. Technical and legal aspects of information protection. Proceedings of the interuniversity student scientific and practical conference], Volgograd, Russia, pp. 189–193.

Hamad, N. and Derar, E. (2022), "Digital Forensics Tools Used in Cybercrime Investigation", *Comparative Analysis. Journal of Xi'an University of Architecture & Technology*, vol. 23, issue 3, pp. 367–379.

Rusetskaya, I.A. (2022), "The role of profiling in ensuring information security", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 3, pp. 85–95.

Rytov, M.Yu. and Kalashnikov, R.Yu. (2019), "Data analysis system for investigating information security incidents in an enterprise network", *Information technologies and systems: management, economics, transport, law*, vol. 2 (34), pp. 208–211.

Shevchenko, D.N. (2020), "Methods of digital forensics and computer forensics for investigating information security incidents – a review", *Problems of science. Computer and Information Sciences*, vol. 4 (52), pp. 32–33.

Smushkin, A.B. (2023), "Digital transformation of forensic science. Part 2 (Digital transformation of the organization of crime investigation, forensic tactics and methods for investigating certain types of crimes)", *Bulletin of the State Law Academy,* no. 2 (151). pp. 272–278.

Yakovleva, A.O. (2023), "On the use of tools for automating the process of incident investigation", *Current issues of aviation and astronautics*, vol. 2, pp. 406–407.

Информация об авторе

Наталия В. Гришина, кандидат технических наук, доцент, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6;

Московский государственный лингвистический университет, Москва, Россия; 119034, Россия, Москва, ул. Остоженка, д. 38 стр. 1; grnat@rambler.ru

Information about the author

Natalia V. Grishina, Cand.of Sci. (Computer Science), associate professor, Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Sq., Moscow, 125047, Russia;

Moscow State Linguistic University, Moscow, Russia; bld. 38-1, Ostozhenka Str., Moscow, 119034, Russia; grnat@rambler.ru

УДК 519.95:004
DOI: 10.28995/2686-679X-2024-2-83-98

О закономерностях при обнаружении атак «отказ в обслуживании» в компьютерных сетях

Николай А. Игнатьев
Национальный Университет Узбекистана им. Мирзо Улугбека, Ташкент, Узбекистан, n_ignatev@rambler.ru

Эркин Р. Наврузов
Национальный Университет Узбекистана им. Мирзо Улугбека, Ташкент, Узбекистан, erkinbek0989@gmail.com

Аннотация. Рассматриваются особенности обнаружения типов DDOS атак в компьютерных сетях. Предложена методика анализа свойств граничных по заданной метрике объектов между типами DDOS атак и нормальным трафиком. Методика ориентирована на поиск закономерностей, связанных с отбором наборов информативных признаков, вычислением значений локальной плотности распределения по граничным объектам классов. Для отбора информативных признаков предложено использовать свойство устойчивости по каждому признаку. Показатель устойчивости не зависит от шкал и масштабов измерений, а также отличается постоянством значений на выборках из генеральной совокупности. Локальные области в форме гипершаров для вычисления плотности распределения представлены объектами одного из классов. Такое представительство связано с выбором радиуса гипершара как расстояния от граничного объекта до первого ближайшего объекта из дополнения к его классу. Значения плотности в гипершаре применяются при анализе отношений объектов класса к граничному. Для интерпретации отношений используются лингвистические переменные и визуальное представление. В форме вычислительного эксперимента демонстрируется связь между отбором информативных признаков и значениями локальной плотности распределения. Применение методики минимизирует затраты вычислительных ресурсов. Решается проблема «проклятия размерности» и снижения вероятности переобучения при машинном обучении.

Ключевые слова: DDOS атаки, информативные признаки, база прецедентов, большие данные, граничные объекты, плотность распределения

Для цитирования: Игнатьев Н.А., Наврузов Э.Р. О закономерностях при обнаружении атак «отказ в обслуживании» в компьютерных сетях // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 83–98. DOI: 10.28995/2686-679X-2024-2-83-98

On patterns in detecting denial of service attacks in computer networks

Nikolai A. Ignat'ev
National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan, n_ignatev@rambler.ru

Erkin R. Navruzov
National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan, erkinbek0989@gmail.com

Abstract. The issues of detecting types of DDOS attacks in computer networks are considered. The methodology is proposed for analyzing the properties of the boundary objects between types of DDOS attacks and normal traffic according to a given metric. A methodology is focused on searching for patterns associated with the selection of sets of informative features and the calculation of local distribution density values for boundary objects of classes. To select informative features, it is proposed to use the stability property for each feature. The stability indicator does not depend on the scales and magnitudes of measurement, and is also distinguished by the constancy of values in selection from the general assembly. Local areas in the form of hyperspheres for calculating the distribution density are represented by objects of one of the classes. Such representation is associated with the choice of the radius of the hypershare as the distance from the boundary object to the first closest object from the complement to its class. The density values in the hypershare are used in the analysis of the relations of objects of the class to the boundary one. Linguistic variables and visual representation are used to interpret relationships. In the form of a computational experiment, the connection between the selection of informative features and the values of the local distribution density is demonstrated. The use of the methodology minimizes the cost of computing resources. The "curse of dimensionality" issue is solved as well as the reducing the probability of overfitting in machine learning is.

Keywords: DDOS attacks, informative futures, precedent base, big data, boundary objects, distribution density

For citation: Ignat'ev, N.A. and Navruzov, E.R. (2024), "On patterns in detecting denial of service attacks in computer networks", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 2, pp. 83–98, DOI: 10.28995/2686-679X-2024-2-83-98

Введение

Исследование особенностей проведения атак «отказ в обслуживании» рассматривается в качестве эффективного средства для защиты компьютерных сетей от несанкционированного доступа. Существенные трудности в процессе исследования связаны с проблемами больших данных (Big Data). Как правило, для выявления скрытых закономерностей или знаний в слабо структурированных предметных областях пользуются методами интеллектуального анализа данных (ИАД) [Алексеев 2020].

Построение эффективных математических моделей, основанных на знаниях, в области информационной безопасности затруднено из-за высокой комбинаторной сложности алгоритмов, наличия проблемы проклятия размерности и формирования баз прецедентов для машинного обучения [Алексеев 2020]. При разделении нормального трафика от атак «отказ в обслуживании» с учетом проблем Big Data требуется использовать алгоритмы с высокой обобщающей способностью и низкими затратами вычислительных ресурсов.

Создать математическую модель, адекватную реальной, возможно за счет использования новых методов, позволяющих выделять наборы информативных признаков и формировать базу прецедентов для машинного обучения. Применительно к информационной безопасности к числу таких методов можно отнести:

– вычисление обобщенных оценок объектов;
– многокритериальный отбор информативных разнотипных признаков;
– редукцию размерности пространства за счет использования свойства устойчивости признаков.

Применение перечисленных методов позволило решить задачи:

– анализа структуры отношений типов DDOS атак;
– вычисления оценок сложности обнаружения типов DDOS атак [Игнатьев, Наврузов 2022];
– отбора информативных признаков для метрических алгоритмов распознавания.

К числу нерешенных проблем можно отнести анализ структуры отношений объектов DDOS атак и нормального трафика. Актуаль-

ность решения этой проблемы сильно возросла после создания методики формирования базы прецедентов [Наврузов 2022]. Привлекательной выглядит идея использования для этих целей редукции размерности пространства для получения визуальной картины отношений объектов.

Искажение структуры данных при проецировании в пространство меньшей размерности связано с изменением отношений близости между объектами. Ранее были предложены два способа для сравнения структур данных до и после проецирования их на двумерную плоскость [Игнатьев, Лолаев 2021]. В этих способах используются сравнение числа точек до и после проецирования и анализ сохранения отношений соседства в исходном и двумерном пространствах.

Существенное значение для формирования структуры отношений между объектами имеет выбор способа нормирования данных. Одной из целей нормирования является инвариантность к масштабам измерений признаков. Свойство инвариантности расширяет возможности для обнаружения скрытых закономерностей, характерных для всех выборок данных из генеральной совокупности.

В настоящей статье нет возможности подробно объяснить особенности реализации процесса многокритериального отбора информативных признаков. Идея этого процесса кратко отображается в виде следующей схемы:

$$<\text{сырые признаки}> \Rightarrow <\text{латентные признаки}> \Rightarrow$$
$$\Rightarrow <\text{информативные признаки}>.$$

Наборы латентных признаков формируются методом обобщенных оценок.

Информативные признаки входят в состав латентного, мера компактности объектов классов по которому на числовой оси максимальна. Мера компактности определяется по максимуму произведения внутриклассового сходства и межклассового различия. Значение меры служит оценкой сложности обнаружения сетевой атаки [Игнатьев, Наврузов 2022]. Отбор информативных признаков производился по парам классов (тип DDOS атаки, нормальный трафик). Количество признаков в информативных наборах по 12 типам DDOS атак было от 3 до 16.

Процессу визуализации [Зиновьев 2000] предшествует селекция обучающей выборки. Проблемы Big Data не позволяют проводить визуализацию DDOS атак по следующим причинам:

– отношения между объектами становятся размытыми из-за «проклятия размерности»;

– описания DDOS атак маскируются под нормальный трафик;
– линейные и нелинейные методы визуализации, такие как PCA и TSNE[1], не отражают структуру отношений между DDOS атаками и нормальным трафиком;
– в нелинейном методе TSNE основным принципом понижения размерности пространства является сохранение плотности распределения в окрестности объектов. Объекты, находящиеся близко друг к другу в пространстве R^n ($n > 3$), при отображении в пространство R^2 таковыми не являются.

Идея предлагаемого метода заключается в анализе плотности распределения в окрестности граничного по заданной метрике объекта. Отличие его от других известных методов вычисления плотности распределения в локальных областях признакового пространства заключается в следующем:

1. Для граничных объектов существует направление в признаковом пространстве по отношению к объектам из противоположных классов.

2. Локальная область данных в форме гипершара с центром в граничном объекте содержит представителей одного класса.

3. Для принятия решений используются результаты анализа отношений объектов одного с граничным объектом класса.

4. Существует количественная оценка отношений объектов, которую можно интерпретировать как визуально, так и с помощью лингвистических переменных.

Вычисление границ интервалов для всех количественных показателей производится рекурсивным алгоритмом по специальному критерию и рассматривается как предобработка данных [Игнатьев, Наврузов 2022]. Границы интервалов и их количество изначально неизвестны и определяются алгоритмическим путем. Целью предобработки является упорядочение признаков по показателю их устойчивости в (0,5;1]. Удаление признаков с относительно малым значением устойчивости направлено на решение проблемы «проклятия размерности» при машинном обучении по выборкам данных.

Постановка задачи

Рассматривается множество (выборка) объектов $E_0 = \{S_1,..., S_m\}$, разбитое на $l + 1$ непересекающихся подмножеств (классов) K_0,

[1] User guide released 0.21.3, 2019 // Scikitlearn. URL: https://scikit-learn.org/ (дата обращения 17.11.2023).

$K_1,\ldots, K_l$, из которых $K_1,\ldots, K_l$ представляют описание l типов DDOS атак, K_0 – нормальной трафик. Объекты выборки задаются набором из n разнотипных признаков $X(n) = (x_1,\ldots, x_n)$, ξ из них измеряются в интервальных шкалах, $n - \xi$ в номинальной шкале. Считается, что определены процедуры для:

– отбора наборов информативных признаков $X(n(j)) \subset X(n)$, $n(j) < n$, по парам (K_0, K_j), $j = 1,\ldots,l$;
– формирования множества $B(j, \rho)$ граничных по заданной метрике $\rho(x, y)$ объектов классов по (K_0, K_j);
– вычисления плотности распределения $F(S, j, \rho)$ по каждому $S \in B(j, \rho)$.

Требуется:

– сформировать множество $B(j, \rho) \subset K_0 \cup K_j$ и вычислить $F(S, j, \rho)$ по каждому $S \in B(j, \rho)$;
– определить наборы информативных признаков $X(n(j))$ по K_0, K_j;
– по $\{F(S, j, \rho)\}$ выделить границы непересекающихся интервалов для лингвистических переменных.

Отбор информативных признаков

Предлагаемый ниже критерий применяется для анализа многообразия отношений значений количественных признаков объектов на числовой оси и предобработки данных [Ignatev, Navruzov 2022]. Особенности предобработки данных заключаются в нелинейных преобразованиях разнотипных (качественных и номинальных) признаков через значения функции принадлежности объектов к классам. Поиску экстремумов критерия предшествует упорядочение значений признаков по неубыванию.

Пусть для значений признака $x_c \in X(n)$ в описании объектов $\Omega_j = K_0 \cup K_j$, $j = 1,\ldots,l$ построена упорядоченная по неубыванию последовательность

$$r_1,\ldots, r_a,\ldots, r_h,\ h = |K_0 \cup K_j|. \qquad (1)$$

Преобразование количественных признаков в градации номинальных по (1) позволяет упростить поиск схожих объектов по обучающей выборке Ω_j с помощью критерия

$$\left|\frac{d_0^i(u,v)}{|K_0|}-\frac{d_j^i(u,v)}{|K_j|}\right|\to\max, \qquad (2)$$

где $d^i_0(u, v)$, $(d^i_j(u, v))$ – количество объектов класса $K_0(K_j)$ в интервале $[r_u; r_v]^i$, $i = 1,\ldots, \tau_c$, τ_c – число непересекающихся интервалов (градаций признака) $x_c \in X(n)$.

Заменим значения признака $x_c \in X(n)$ у объектов из Ω_j на номера интервалов $[r_u; r_v]^\mu$, $\mu = 1,\ldots,\tau_c$. Тогда $d_{0c}(\mu) = d^\mu_0(u, v)$, $d_{jc}(\mu) = d^\mu_j(u, v)$. Значение функции принадлежности признака $x_c \in X(n)$ к K_0 определяется как

$$f_c(\mu)=\frac{d_{0c}(\mu)/|K_0|}{d_{0c}(\mu)/|K_0|+d_{jc}(\mu)/|K_j|}. \qquad (3)$$

Обозначим $z_c(\mu) = d_{0c}(\mu) + d_{jc}(\mu)$. Тогда вычисление значения устойчивости признака $x_c \in X(n)$ по (3) будет таким:

$$g_c=\frac{1}{m}\sum_{\mu=1}^{\tau}\begin{cases} f_c(\mu)\cdot z_c(\mu), f_c(\mu)>0.5,\\ (1-f_c(\mu))\cdot z_c(\mu), f_c(\mu)<0.5\end{cases}. \qquad (4)$$

Заметим, что вариант с $f_c(\mu) = 0{,}5$ в (4) не рассматривается. Считается, что вероятность такого события в задачах с Big Data близка к нулю.

Пусть

$$x^1,\ldots, x^n, \qquad (5)$$

упорядоченная по невозрастанию устойчивости (4) последовательность признаков. Для формирования информативного набора предлагается эвристика, суть которой заключается в последовательном удалении из (5) признаков с малым значением устойчивости. Как правило, мощность набора определяется по результатам вычислительного эксперимента. К числу важнейших показателей эксперимента относят обобщающую способность алгоритмов распознавания.

Анализ отношений по граничным объектам

Формирование признакового пространства является важным этапом решения задач классификации. Специфичность рассматриваемой задачи формирования пространства состоит в том, что сырые признаки представлены в разных шкалах и масштабах измерений [Игнатьев, Лолаев 2021].

Для анализа особенностей типов DDOS атак предлагается использовать значения плотности распределения по локальным областям в форме гипершаров. Центрами гипершаров являются граничные по заданной метрике объекты классов. Специфика использования гипершаров выражается в:

– выборе радиуса с центром в граничном объекте;
– оценке плотности распределения объектов в гипершаре и ее интерпретации.

Значение радиуса r_s определяется как расстояние от граничного объекта $S \in K_t$ до ближайшего объекта $S' \in CK_t$, $t = 1,\ldots,l$. С помощью радиуса r_s формируется множество объектов $\{S_i \in K_t \mid \rho(S, S_i) < r_s\}$. Плотность распределения в гипершаре относительно объекта S определяется как

$$zich(S, r_s) = \sum_{\rho(S,S_i)<r_s} \left(1 - \frac{d_i}{r_s}\right) / k \,, \qquad (6)$$

где $d_i = \rho(S, S_i)$, k – число объектов в гипершаре, включая объект S. Очевидно, что значение плотности (6) варьируется в границах (0;1].

Введение лингвистических переменных позволяет интерпретировать смысл отношений объектов в гиперпространстве на основе лишь значений плотности и числа объектов в гипершарах. Значения лингвистической переменной предлагается определять (см. рис. 1, 2, 3, 4) в зависимости от принадлежности плотности (6) к следующим интервалам:

- (0; 0,35] – не уважают;
- (0,35; 0,75] – уважение нормальное;
- (0,75; 1) – очень уважают;
- [1; 1] – себя уважает.

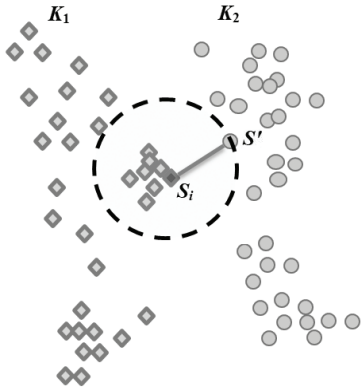


Рис. 1. Отношение «очень уважают»

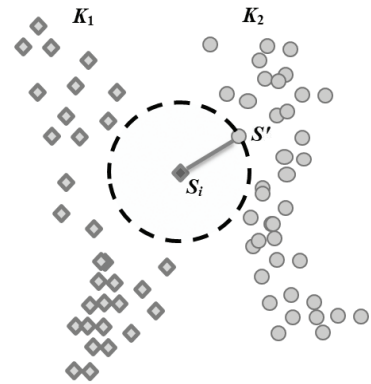


Рис. 2. Отношение «себя уважает»

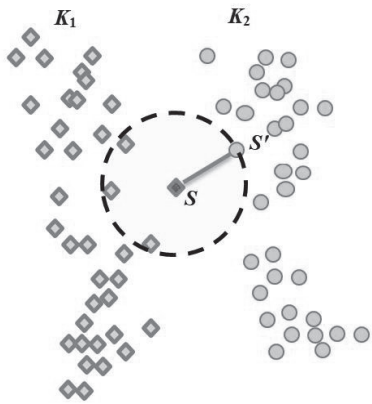


Рис. 3. Отношение «не уважают»

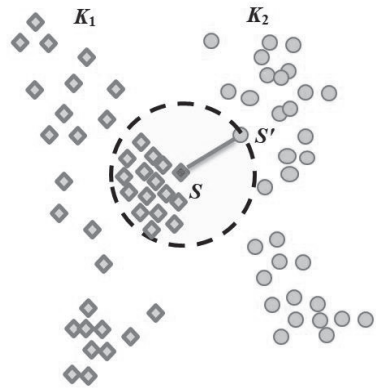


Рис. 4. Отношение «уважение нормальное»

Вычислительный эксперимент

База прецедентов для эксперимента была представлена 2000 объектов, описываемыми 80 признаками, из них в K_1 (DDOS атаки типа DNS) – 1000, в K_2 (нормальный трафик) – 1000 объектов.

Примеры интерпретации специфики DDOS атак типа DNS по евклидовой метрике в (6) с помощью лингвистических переменных приводятся в табл. 1.

Таблица 1

Примеры отношений
к граничным объектам DDOS атак типа DNS

№ п/п	Мощность гипершара	Значение	
		плотности	лингвистической переменной
1	1	1,0	себя уважает
2	2	0,8615	очень уважают
3	7	0,6752	уважение нормальное
4	4	0,3723	уважение нормальное
5	4	0,7599	очень уважают
6	1	1,0	себя уважает
7	3	0,4400	уважение нормальное
8	2	0,7019	уважение нормальное
9	5	0,6359	уважение нормальное
10	2	0,8465	очень уважают
11	2	0,9516	очень уважают
12	4	0,3045	не уважают
13	4	0,3723	уважение нормальное
14	2	0,5395	уважение нормальное
15	6	0,3304	не уважают

Укажем выводы, которые можно сделать (см. табл. 1) на основе значений плотности (6) для граничных объектов.

1. Для состава гипершаров нехарактерна большая мощность.

2. Значение плотности, равное 1, указывает на наличие изолированных объектов из состава атакующей стороны.

Одной из целей эксперимента является демонстрация эффективности отбора информативных наборов признаков на основе свойства их устойчивости. Доказательство эффективности предложено строить с помощью упорядоченной по невозрастанию значений устойчивости последовательности признаков. Удаление признаков с относительно малым значением устойчивости отслеживается через изменение мощности множества граничных объектов, значение плотности (6) в их окрестности и визуализацию объектов по методу TSNE. В табл. 2 представлена упорядоченная по значениям устойчивости признаков (4) последовательность для DDOS атак типа DNS.

Таблица 2

Упорядоченная по значениям устойчивости последовательность признаков для DDOS атак типа DNS

№ п/п	Номер признака	Значение устойчивости	Название признаков
1	39	0,9711	Минимальная длина пакета
2	8	0,9707	Минимальный размер пакета в прямом направлении
3	15	0,9671	Количество потоковых пакетов в секунду
4	9	0,9440	Средний размер пакета в прямом направлении
5	54	0,9440	Средний размер наблюдается в прямом направлении
6	53	0,9233	Средний размер пакета
7	70	0,9211	Минимальный размер сегмента наблюдается в прямом направлении
8	4	0,9206	Всего пакетов в обратном направлении
...	...	...	...
31	42	0,8156	Стандартное отклонение длины пакета
32	43	0,8156	Длина отклонения пакета
33	11	0,7676	Максимальный размер пакета в обратном направлении
34	13	0,7676	Средний размер пакета в обратном направлении
35	66	0,7676	Среднее количество байтов в подпотоке в обратном направлении
36	55	0,7676	Средний объем байтов в прямом направлении
37	6	0,7676	Общий размер пакета в обратном направлении
38	67	0,7483	Общее количество байтов, отправленных в начальном окне в прямом направлении
39	22	0,7418	Среднее время между двумя пакетами, отправленными в прямом направлении

Окончание табл. 2

№ п/п	Номер признака	Значение устойчивости	Название признаков
40	20	0,7362	Минимальное время между двумя пакетами, отправляемыми в потоке
...	...	...	...
62	71	0,5515	Среднее время, когда поток был активен, прежде чем стал свободным
63	46	0,5430	Количество пакетов с RST
64	31	0,5430	Количество раз, когда флаг PSH был установлен в пакетах, движущихся в прямом направлении (0 для UDP)
65	76	0,5372	Стандартное отклонение времени, в течение которого поток простаивал перед активацией
66	72	0,5362	Стандартное отклонение времени, в течение которого поток был активен до простоя
68	32	0	Количество раз, когда флаг PSH был установлен в пакетах, движущихся в обратном направлении (0 для UDP)
69	57	0	Средний объем байтов в прямом направлении
...	...	...	...
71	59	0	Среднее количество насыпного курса в прямом направлении
76	33	0	Количество раз, когда флаг URG был установлен в пакетах, проходящих в прямом направлении (0 для UDP)
...	...	...	...
79	79	0	Подобный HTTP
80	44	0	Количество пакетов с FIN

Анализ результатов из табл. 2 показывает, что начиная с 62 элемента последовательности признаки следует отнести к числу малоинформативных, а с 68 – к числу неинформативных. Имеет смысл проверка утверждения, что удаление признаков с низким значением

устойчивости не вносит существенных изменений в структуру отношений объектов. Результаты проверки данного утверждения по нескольким упорядоченным наборам признаков (см. табл. 1) демонстрируются в табл. 3.

Таблица 3

Анализ свойств граничных объектов DDOS атак типа DNS по разным наборам упорядоченных признаков

№ п/п	Количество признаков					
	80		40		30	
	Значение плотности	Мощность гипершара	Значение плотности	Мощность гипершара	Значение плотности	Мощность гипершара
1	0,944	19	0,944	19	0,947	19
2	0,500	2	0,450	2	0,450	2
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	0,944	19	0,944	19	1	1
6	1	1	1	1	1	1
7	1	1	1	1	0,499	2
8	0,500	2	0,450	2	0,269	3
9	0,270	3	0,269	3	0,799	54
10	0,799	54	0,799	54	0,799	54
11	0,799	54	0,799	54	0,295	3
12	0,295	3	0,295	3	0,146	3
13	0,146	3	0,146	3	0,715	54
14	0,715	54	0,715	54	0,167	2
15	0,166	2	0,168	2	0,341	14
16	0,715	22	–	–	–	–

Незначительные изменения состава граничных объектов и значений плотности в их окрестности доказывают эффективность использования показателей устойчивости для отбора информативных признаков.

Другое подтверждение результата из табл. 3 показано на рис. 5 и 6 через визуальное представление свойств граничных объектов DDOS атак типа DNS по разным наборам упорядоченных признаков с помощью метода TSNE в R^2.

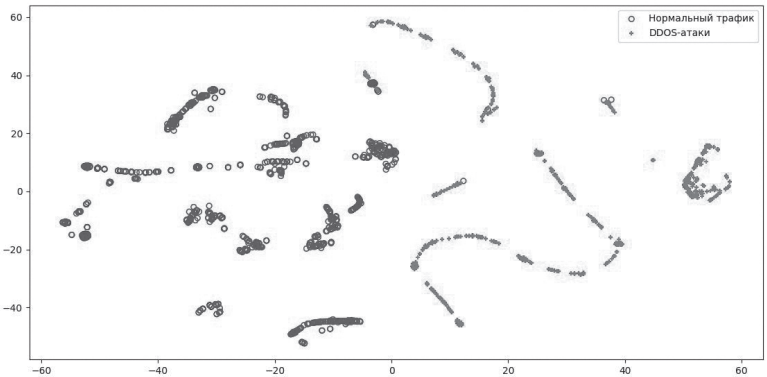


Рис. 5. Визуализация DDOS атак типа DNS по всем признакам

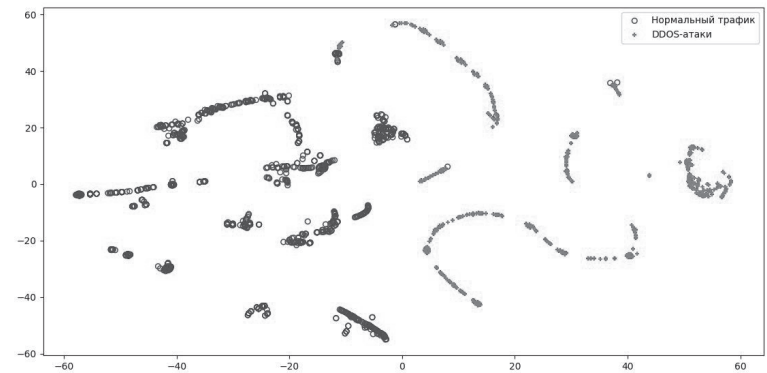


Рис. 6. Визуализация DDOS атак типа DNS по 40 признакам из упорядоченной по значениям устойчивости последовательности

На рис. 5 и 6 видно, что при удалении неинформативных признаков структура отношений объектов сильно не меняется.

Заключение

Вычисление значений плотности распределения по гипершарам с центрами в граничных объектах классов дает знание о наличии скрытых закономерностей в данных. Характерной закономерностью является отсутствие ярко выраженных кластерных структур, указывающих на наличие компактного размещения в признаковом пространстве представителей атакующих и нормального трафика. Открываются возможности для объяснения причин сложности обнаружения типов DDOS атак.

Литература

Алексеев 2020 – *Алексеев И.В.* Обнаружение распределенных атак отказа в обслуживании в крупномасштабных сетях на основе методов математической статистики и искусственного интеллекта: Автореф. дис. ... канд. техн. наук. СПб., 2020. 20 с.

Игнатьев, Лолаев 2021 – *Игнатьев Н.А., Лолаев М.Я.* Анализ соответствия структур отношений объектов классов на многообразиях их описаний // Информационные Технологии. 2021. Т. 27. № 1. С. 18–24. DOI: 10.17587/it.27.18-24.

Ignatev, Navruzov 2022 – *Ignatev N.A., Navruzov E.R.* Estimates of the complexity of detecting types of DDOS attacks // International Journal of Computing. 2022. Vol. 21, issue 4. P. 443–449. DOI: 10.47839/ijc.21.4.2779.

Наврузов 2022 – *Наврузов Э.Р.* О формировании баз прецедентов для решения задач информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2022. № 3. С. 66–84. DOI: 10.28995/2686-679X-2022-3-66-84.

Зиновьев 2000 – *Зиновьев А. Ю.* Визуализация многомерных данных. Красноярск, КГТУ, 2000. 180 с.

References

Alekseev, I.V. (2020), *Obnaruzhenie raspredelennykh atak otkaza v obsluzhivanii v krupnomasshtabnykh setyakh na osnove metodov matematicheskoi statistiki i iskusstvennogo intellekta* [Detection of distributed denial of service attacks in large-scale networks based on mathematical statistics and Artificial Intelligence methods], Abstract of Ph.D. Thesis, St. Petersburg, Russia.

Ignat'ev, N.A. and Lolaev, M.Ya. (2021), "Analysis in the correspondence of the relations structures of the class objects on the varieties of their descriptions", *Information Technologies*, vol. 27, no. 1, pp. 18–24, DOI: 10.17587/it.27.18-24.

Ignat'ev, N.A. and Navruzov, E.R. (2022), "Estimates of the complexity of detecting types of DDOS attacks", *International Journal of Computing*, vol. 21, issue 4, pp. 443–449, DOI: 10.47839/ijc.21.4.2779.

Navruzov, E.R. (2022), "On forming the precedent bases for solving problems of the information security", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 3, pp. 66–84, DOI: 10.28995/2686-679X-2022-3-66-84.

Zinov'ev, A.Yu. (2000), *Vizualizatsiya mnogomernykh dannykh* [Visualization of multidimensional data], KSTU, Krasnoyarsk, Russia.

Информация об авторах

Николай А. Игнатьев, доктор физико-математических наук, профессор, Национальный Университет Узбекистана им. Мирзо Улугбека, Ташкент, Узбекистан; 100174, Республика Узбекистан, Ташкент, ул. Университетская, д. 4; n_ignatev@rambler.ru

Эркин Р. Наврузов, старший преподаватель, Национальный Университет Узбекистана им. Мирзо Улугбека, Ташкент, Узбекистан; 100174, Республика Узбекистан, Ташкент, ул. Университетская, д. 4; erkinbek0989@gmail.com

Information about the authors

Nikolai A. Ignat'ev, Dr. of Sci. (Physics and Mathematics), professor, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan; bld. 4, Universitetskaya Str., Tashkent, 100174, Uzbekistan; n_ignatev@rambler.ru

Erkin R. Navruzov, senior lecturer, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan; bld. 4, Universitetskaya Str., Tashkent, 100174, Uzbekistan; erkinbek0989@gmail.com

УДК 681.326+519.87+004
DOI: 10.28995/2686-679X-2024-2-99-108

Информационное обеспечение импульсного управления устойчивостью систем информационной безопасности

Андрей С. Кузнецов
Российский государственный социальный университет, Москва, Россия, askgoogle@internet.ru

Андрей Е. Краснов
Российский государственный социальный университет, Москва, Россия, krasnovmgutu@yandex.ru

Аннотация. В статье подробно рассмотрены основные этапы создания информационного обеспечения процесса импульсного управления для обеспечения устойчивости систем информационной безопасности. Проведен структурный системный анализ процессов и моделирование импульсного управления устойчивостью автоматизированных систем информационной безопасности. Выполнено формализованное описание процесса управления выбором оптимальной математической модели сохранения для показателя устойчивости автоматизированных информационных систем на примере информационных систем управления информационной безопасностью. Приведен алгоритм высокоэффективного управления устойчивостью системы информационной безопасности. Построено информационное описание для оптимизации управления процессами выбора модели импульсного управления на основе комплекса функциональных моделей, детализирующих и формализующих процессы управления устойчивостью автоматизированных систем информационной безопасности. Реализованы основные принципы стратегического управления для показателя устойчивости автоматизированных систем информационной безопасности на основе критериального подхода. Рассмотрена концепция модели импульсного управления на основе регулятора для поддержания динамической устойчивости функционирования автоматизированных систем информационной безопасности. Рассмотренное алгоритмическое и информационное обеспечение представляет собой динамическую модель интеллектуальной информационной поддержки процесса принятия решений по выбору оптимальной модели, реализующей принцип импульсного управления в автоматизированных системах обеспечения информационной безопасности.

Ключевые слова: модель управления, алгоритмическое обеспечение, информационная поддержка, импульсное управление, устойчивость систем информационной безопасности

Для цитирования: Кузнецов А.С., Краснов А.Е. Информационное обеспечение импульсного управления устойчивостью систем информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 2. С. 99–108. DOI: 10.28995/2686-679X-2024-2-99-108

Information support for pulse control of the stability of information security systems

Andrei S. Kuznetsov
Russian State Social University, Moscow, Russia,
askgoogle@internet.ru

Andrei E. Krasnov
Russian State Social University, Moscow, Russia,
krasnovmgutu@yandex.ru

Abstract. The article considers in detail the main stages of creating information support for the impulse control process to ensure the stability of information security systems. The authors carried out a structural system analysis of processes and modeling of impulse control of the stability in automated information security systems. Also a formalized description of the process of managing the selection of the optimal mathematical conservation model for the stability indicator of automated information systems is carried out using the example of information security management information systems. The article presents an algorithm for highly effective management of the stability of an information security system. An information description is constructed to optimize the management of processes for selecting an impulse control model based on a set of functional models that detail and formalize the processes of managing the stability of automated information security systems. The basic principles of strategic management for the sustainability indicator of automated information security systems are implemented based on a criteria-based approach. The concept of a controller-based impulse control model for maintaining the dynamic stability of the functioning of automated information security systems is considered. The considered algorithmic and information support is a dynamic model providing intelligent information support of the decision-making process for choosing the optimal model realizing the principle of impulse control in automated information security systems.

Keywords: control model, algorithmic support, information support, impulse control, stability of information security systems

For citation: Kuznetsov, A.S. and Krasnov, A.E. (2024), "Information support for pulse control of the stability of information security systems", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 2, pp. 99–108, DOI: 10.28995/2686-679X-2024-2-99-108

Введение

В настоящее время проблема управления устойчивостью автоматизированных систем обеспечения информационной безопасности является достаточно актуальной задачей. Основные пути оптимизации решения задачи управления в системах обеспечения информационной безопасности подробно изложены в литературе [Соколов 2007; Тутубалин 2017; Гавдан 2022; Краснов 2021; Арутюнов 2023; Надеждин 2009; Надеждин 2012; Краснов 2023]. Модели анализа и обеспечения устойчивости могут быть интерпретированы как:

1) устойчивое обеспечение информационной безопасности с применением определенных средств защиты [Соколов 2007];
2) как подход к обеспечению безопасности и устойчивости функционирования автоматизированной информационной системы на основе набора принципов безопасного устойчивого функционирования [Тутубалин 2017];
3) устойчивое обеспечение функционирования информационной системы построено на вычислении устойчивости ее активов и информационных технологий рекуррентного пересчета при добавлении в систему новых компонентов [Гавдан 2022];
4) алгоритмы поиска наиболее уязвимых элементов технологической системы и связей между работоспособностью этих элементов и защищенностью информационных потоков в системах управления технологическими процессами [Краснов 2021].

Одним из направлений среди возможных способов обеспечения информационной безопасности автоматизированных информационных систем является применение импульсной модели управления. В общем виде модель импульсного управления можно описать как операции по вложению денежных средств, представленные в виде периодов поддержания или настройки устойчивости автоматизированных систем информационной безопасности (СИБ).

В качестве модели потери первоначальной устойчивости СИБ выбрана экспоненциальная модель (описание в моменты времени n дискретного времени $t_n = n\Delta t$.):

$$X_n \quad \frac{1}{1+\frac{\Delta t}{T}} X_{n-1};\ X_{n=0} = X_0;\ n = 1, 2, \ldots, 13, \qquad (1)$$

где Δt – интервал времени, а T – период потери устойчивости [Краснов 2024].

Для описания процессов управления устойчивостью автоматизированных систем информационной безопасности приведена их формализация – разработан комплекс функциональных моделей, а также приведена схема программы реализации импульсной модели управления. Информационное описание в виде функциональных схем позволяет детализировать операции по управлению устойчивостью информационных систем управления информационной безопасностью.

На рис. 1 приведена обобщенная функциональная диаграмма уровня А-0, отражающая организацию процесса управления в автоматизированных системах обеспечения информационной безопасности в первом приближении.

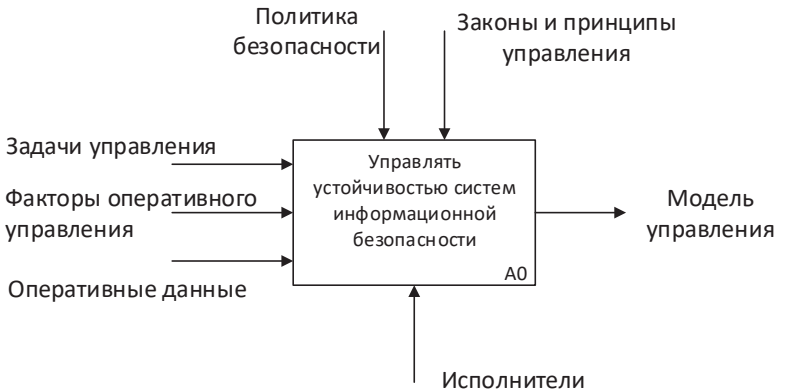


Рис. 1. Диаграмма уровня А-0 процесса управления устойчивостью систем информационной безопасности

В данной контекстной диаграмме входными потоками являются: поставленные задачи по управлению устойчивостью автоматизированных систем информационной безопасности, оперативные данные (информация) по динамике изменения устойчивости

систем информационной безопасности. Потоки управления представлены законами, основными принципами управления, а также политиками информационной безопасности, включая механизмы управления рисками. Стрелка типа механизм реализуется отдельным исполнителем – разработчиком импульсной модели управления устойчивостью автоматизированных систем информационной безопасности. На выходе представлена разработанная модель управления. Основные факторы оперативного управления представлены на рис. 2.

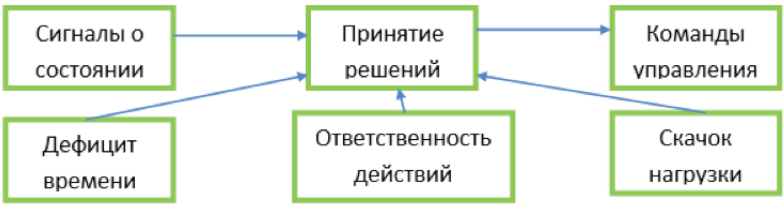


Рис. 2. Факторы оперативного управления
в процессах принятия управленческих решений

Далее была выполнена функциональная декомпозиция обобщенной функциональной диаграммы уровня А-0 и получена дочерняя диаграмма уровня А-1, позволяющая выполнить детализацию описания процессов разработки модели импульсного управления устойчивостью автоматизированных систем информационной безопасности на основе принципа управления по отклонению. Диаграмма представляет собой дальнейшую детализацию процессов управления устойчивостью автоматизированных систем обеспечения информационной безопасности. Она описывает последовательные преобразования от исходных данных и сформулированных задач и принципов управления к созданию модели импульсного управления в системах информационной безопасности на основе регламентирующих документов (техническая документация (ТД), регламентирующие документы (РД)) и действий индивидуальных исполнителей на каждой конкретной стадии процесса (инженер по информационной безопасности (ИБ)). На выходе мы имеем конкретизированную модель высокоэффективного управления и параметры оптимального управления (рис. 3).

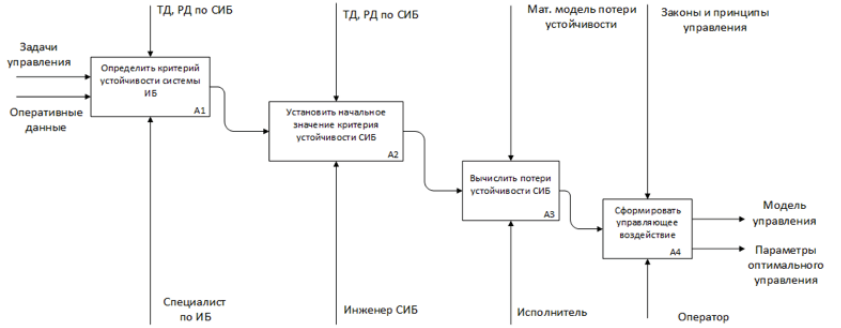


Рис. 3. Диаграмма уровня А-1.
Синтез импульсной модели управления процессами обеспечения устойчивости систем информационной безопасности

На рис. 4 приведена схема программы управления устойчивостью автоматизированных систем информационной безопасности на основе импульсной модели. На начальном этапе происходит определение критерия устойчивости и установка его начального значения. Далее выполняется формализация процесса управления на основе выбранной математической модели. Последовательно рассчитываются значения потерь устойчивости системы информационной безопасности и анализируются причины отклонения показателя устойчивости. Затем формируется импульсное управляющее воздействие для восстановления значений показателя устойчивости автоматизированной системы информационной безопасности. Подробное описание математических моделей потери устойчивости в системах информационной безопасности приведено в статье [Краснов 2024]. При необходимости производится корректировка параметров или вида выбранной математической модели (параметрический или структурный синтез – модификация).

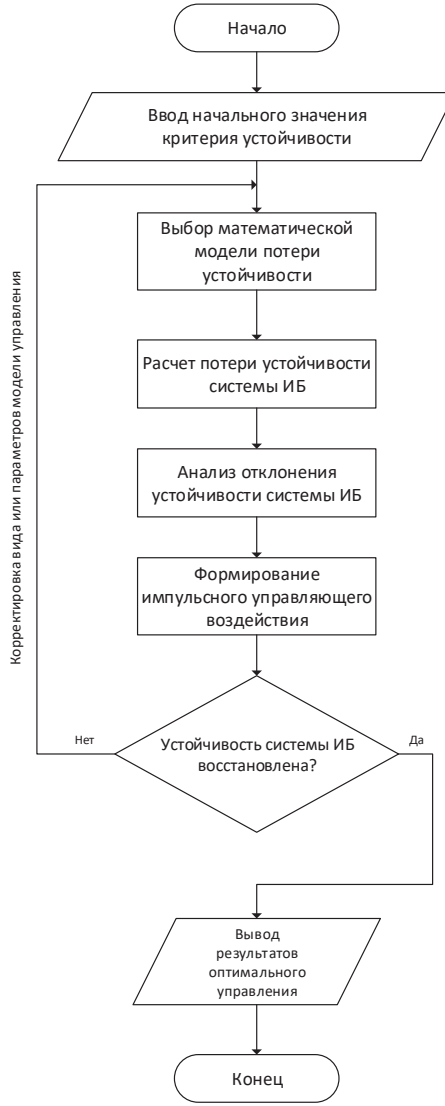


Рис. 4. Схема программы поддержания устойчивости системы информационной безопасности на основе импульсной модели управления

Заключение

Проведен анализ литературных источников, и установлены основные инструменты обеспечения устойчивости в системах информационной безопасности. Выполнена формализация процесса управления выбором оптимальной модели сохранения устойчивости автоматизированных информационных систем на примере информационных систем управления информационной безопасностью. На основе системного подхода реализовано информационное описание процесса выбора модели импульсного управления на основе комплекса функциональных моделей процессов управления устойчивостью автоматизированных систем информационной безопасности. Приведен алгоритм высокоэффективного управления устойчивостью системы информационной безопасности на основе критериального подхода и импульсной модели управления. Предложено информационное обеспечение и схема программы управления в качестве динамической модели информационной поддержки процесса принятия решений по выбору оптимальной модели импульсного управления в автоматизированных системах обеспечения информационной безопасности.

Литература

Арутюнов 2023 – *Арутюнов В.В., Гришина Н.В.* Об итогах VI Всероссийской научно-практической конференции «Информационная безопасность: вчера, сегодня, завтра» // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 3. С. 38–48. DOI: 10.28995/2686-679X-2023-3-38-48.

Гавдан 2022 – *Гавдан Г.П. и др.* Устойчивость функционирования объектов критической информационной инфраструктуры // Безопасность информационных технологий [S.l.]. 2022. Т. 29. № 4. С. 53–66.

Краснов 2021 – *Краснов А.Е., Мосолов А.С., Феоктистова Н.А.* Оценивание устойчивости критических информационных инфраструктур к угрозам информационной безопасности // Безопасность информационных технологий [S.l.]. 2021. Т. 28. № 1. С. 106–120.

Краснов 2023 – *Краснов А.Е., Чеканов И.Р., Козочкин И.Д.* Автоматизация поддержки принятия управленческих решений в области информационной безопасности на основе технологии экспертных систем // Информатизация образования и науки. 2023. № 2 (58). С. 81–89.

Краснов 2024 – *Краснов А.Е., Кузнецов А.С., Смирнов В.М.* Модель импульсного управления устойчивостью системы информационной безопасности // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2024. № 1. С. 80–90. DOI: 10.28995/2686-679X-2024-1-80-90.

Надеждин 2012 – *Надеждин Е.Н.* Проблемные вопросы управления рисками информационной безопасности в сфере образования // Научный поиск. 2012. № 2.6. С. 50–57.

Надеждин 2009 – *Надеждин Е.Н., Смирнова Е.Е., Козлов А.О.* Модели информационного противоборства в задачах оценки безопасности вычислительных сетей. Информатизация образования и науки. 2009. № 2. С. 45–50.

Соколов 2007 – *Соколов Б.В., Охтилев М.Ю.* и др. Методы и алгоритмы оперативного решения задач оценивания показателей возможностей и устойчивости функционирования информационной системы // Труды СПИИРАН. 2007. № 4. С. 255–269.

Тутубалин 2017 – *Тутубалин П.И., Кирпичников А.П.* Модель анализа устойчивого управления информационной безопасностью распределенной информационной системой // Вестник технического университета. 2017. Т. 20. № 19. С. 96–101.

References

Arutyunov, V.V. and Grishina, N.V. (2023), "On the results of the 6th All-Russian Scientific and Practical Conference 'Information Security. Yesterday, Today, Tomorrow' ", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 3. pp. 38–48, DOI: 10.28995/2686-679X-2023-3-38-48.

Gavdan, G.P. et al. (2022), "Stability in the functioning of critical information infrastructure objects", *Security of information technologies, [S.l.]*, vol. 29, no. 4, pp. 53–66.

Krasnov, A.E., Mosolov, A.S. and Feoktistova, N.A. (2021), "Assessing the resilience of critical information infrastructures to information security threats", *Information Technology Security, [S.l.]*, T. 28, no. 1, pp. 106–120.

Krasnov, A.E., Chekanov, I.R. and Kozochkin, I.D. (2023), "Automation of support for management decisions in the field of information security based on expert systems technology", *Informatization of education and science*, no. 2 (58), pp. 81–89.

Krasnov, A.E., Kuznetsov, A.S. and Smirnov, V.M. (2024) "The model of pulse control of the information security system stability", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series*, no. 1. P. 80–90. DOI: 10.28995/2686-679X-2024-1-80-90.

Nadezhdin, E.N. (2012), "Problematic issues of information security risk management in the field of education", *Scientific search*, no. 2.6, pp. 50–57.

Nadezhdin, E.N., Smirnova, E.E. and Kozlov, A.O. (2009), "Models of information warfare in tasks of assessing the security of computer networks", *Informatization of education and science*, no. 2, pp. 45–50.

Sokolov, B.V., Okhtilev, M.Yu. et al. (2007), "Methods and algorithms for quickly solving problems of assessing indicators of capabilities and stability in the functioning of an information system", *Proceedings of SPIIRAS*, no. 4, pp. 255–269.

Tutubalin, P.I. and Kirpichnikov, A.P. (2017), "Model for analyzing sustainable information security management of a distributed information system", *Bulletin of Technical University*, vol. 20, no. 19, pp. 96–101.

Информация об авторах

Андрей С. Кузнецов, кандидат технических наук, доцент, Российский государственный социальный университет, Москва, Россия; 129226, Россия, Москва, ул. В. Пика, д. 4, стр. 1; askgoogle@internet.ru

Андрей Е. Краснов, доктор физико-математических наук, профессор, профессор, Российский государственный социальный университет, Москва, Россия; 129226, Россия, Москва, ул. В. Пика, д. 4, стр. 1; krasnovmgutu@yandex.ru

Information about the authors

Andrei S. Kuznetsov, Cand. of Sci. (Mechanical Engineering), associate professor, Russian State Social University, Moscow, Russia; bld. 4/1, V. Pika Str., Moscow, 129226, Russia; askgoogle@internet.ru

Andrei E. Krasnov, Dr. of Sci. (Physics and Mathematics), professor, Russian State Social University, Moscow, Russia; V. Pika Str., bld. 4/1, Moscow, 129226, Russia; krasnovmgutu@yandex.ru

Научный журнал
Вестник РГГУ
Серия «Информатика.
Информационная безопасность. Математика»
№ 2
2024

Дизайн обложки
Е.В. Амосова

Корректор
А.А. Леонтьева

Компьютерная верстка
Н.В. Москвина

Учредитель и издатель
Российский государственный гуманитарный университет
125047, Москва, Миусская пл., 6

Свидетельство о регистрации СМИ
ПИ № ФС77-72977 от 25.05.2018 г.
Периодичность 4 раза в год

Подписано в печать 11.06.2024
Выход в свет 18.06.2024
Формат 60×90 1/16
Уч.-изд. л. 6,8. Усл. печ. л. 6,9
Тираж 1050 экз. Свободная цена
Заказ № 1994

Отпечатано в типографии Издательского центра
Российского государственного гуманитарного университета
125047, Москва, Миусская пл., 6
www.rsuh.ru