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

1
2023

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. Peer-reviewed publications fall within the following research area:

**20.00.00** Informatics

**81.93.29** Information security, data protection

**27.00.00** Mathematics

*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). Научные рецензируемые публикации соответствуют отраслям науки:

**20.00.00** Информатика

**81.93.29** Информационная безопасность, защита информации

**27.00.00** Математика

*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

СОДЕРЖАНИЕ

УДК 004.67
DOI: 10.28995/2686-679X-2023-1-8-22

# Визуализация графа для анализа неформальных социальных связей в студенческой среде

Дмитрий Г. Алпацкий
*Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, alpackij@bmstu.ru*

Алексей А. Канаев
*Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, alekseykanaev@mail.ru*

Андрей И. Колбин
*Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, kolbinai@student.bmstu.ru*

Дмитрий Д. Ставский
*Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия, stavskiydd@student.bmstu.ru*

*Аннотация.* В связи с ростом потребности в социальной аналитике и увеличением количества возможных параметров, которые подходят для исследования, становятся все более востребованными новые пути применения метода визуализации. Такую потребность могут удовлетворить специальные элементы теории графов, а именно – отрисовка графа в двух- или трехмерном формате. Это может быть реализовано при помощи различных средств, которые представлены современными компьютерными технологиями. Данные, взятые в качестве примера, основываются на выявлении дружеских связей между студентами вуза в социальной сети и обрабатываются при помощи различных математических моделей, часть из которых была разработана авторами. Подчеркивается, что после сбора данные были агрегированы до уровня учебных групп и таким образом деанонимизированы. Интерактивная визуализация отмечается как удобный и быстрый способ проверки гипотез и проведения поисковых исследований на данных социальных сетей. Подбор данных не ограничивается потребностями вуза и может быть расширен до определенного круга

заинтересованных сторон. В связи с этим в статье предложена методика применения подходов к визуализации графа для анализа неформальных социальных (дружеских) связей.

*Ключевые слова:* социальные связи, визуализация графов, агрегация, фильтрация, детализация

*Для цитирования: Алпацкий Д.Г., Канаев А.А., Колбин А.И., Ставский Д.Д.* Визуализация графа для анализа неформальных социальных связей в студенческой среде // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 1. С. 8–22. DOI: 10.28995/2686-679X-2023-1-8-22

# Graph visualization for the analysis of informal social relations in the student surroundings

Dmitrii G. Alpatskii
*Bauman Moscow State Technical University, Moscow, Russia, alpackij@bmstu.ru*

Aleksei A. Kanaev
*Bauman Moscow State Technical University, Moscow, Russia, alekseykanaev@mail.ru*

Andrei I. Kolbin
*Bauman Moscow State Technical University, Moscow, Russia, kolbinai@student.bmstu.ru*

Dmitrii D. Stavskii
*Bauman Moscow State Technical University, Moscow, Russia, stavskiydd@student.bmstu.ru*

*Abstract*. Due to the growing need for social analytics and an increase in the number of possible parameters that are suitable for research, new ways of applying the visualization method are becoming more and more in demand. Such a need can be met by special elements of graph theory, namely, drawing a graph in two- or three-dimensional format. It can be implemented with the help of various means, which are represented by modern computer technologies. The data taken as an example is based on the identification of friendly ties between university students in a social network and is processed using various mathematical models, some of which were developed by the authors. It is emphasized that after the collection, the data were ag-

gregated to the level of study groups and thus deanonymized. Interactive visualization is noted as a convenient and fast way to test hypotheses and conduct search studies on social media data. The selection of data is not limited to the needs of the university and can be expanded to a certain range of stakeholders. In that regard, the article proposes a methodology for applying approaches to graph visualization for the analysis of informal social (friendly) ties.

*Keywords:* social connections, graph visualization, aggregation, filtering, detailing

*For citation:* Alpatskii, D.G., Kanaev, A.A., Kolbin, A.I. and Stavskii, D.D. (2023), "Graph visualization for the analysis of informal social relations in the student surroundings", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 1, pp. 8–22, DOI: 10.28995/2686-679X-2023-1-8-22

## *Введение*

Анализ данных социальных сетей – это современный тренд как в прикладных, так и в фундаментальных науках. Об этом свидетельствуют результаты анализа наукометрических показателей и контент-анализа публикаций: 912 тыс. публикаций по теме, проиндексированных в Google Scholar, 41 тыс. – на Cyberleninka.org.

Открытая общедоступная информация, добровольно публикуемая пользователями социальных сетей, является источником нового знания об общественных связях, которые могут помочь в осуществлении различных видов деятельности: рекламной, управленческой, исследовательской, образовательной, деятельности по обеспечению безопасности.

Социальные сети – относительно легкодоступный источник хорошо структурированных данных в машиночитаемом формате. Формально данные социальной сети сводятся к совокупности узлов, их атрибутов и связей между ними [Мельникова, Яковлев 2014, с. 255]. Заметим, что такая структура данных хорошо описывается графовыми моделями [Долинина, Печенкин, Тарасова 2009]. Использование графических компьютерных технологий позволяет довести визуализацию для дальнейшего практического применения.

Авторы статьи провели исследование подходов к построению графов дружеских связей, полученных из социальных сетей: 2D- и 3D-визуализация; агрегация и фильтрация данных; математическая детализация.

Цель работы – разработать методику применения визуализации графа для анализа неформальных социальных (дружеских) связей студентов в вузе.

## *Актуальность работы*

Информационное пространство современного вуза постоянно расширяется. Неформальные социальные связи между студентами представляют собой достоверную коммуникационную модель их профилей в сети [Соловьев, Гончаров, Батин 2020]. Данный аспект жизни университета недостаточно изучен с использованием современных подходов к визуализации графов.

## *Граф как метод визуализации данных*

В начале представления результатов исследования обозначим, что такое граф и в чем преимущество его использования. Графы – это математические структуры, используемые для изучения парных отношений между объектами [Линник 2018].

Визуализация графа, также называемая анализом связей или визуализацией сети, представляет собой способ визуального представления связей между объектами в данных. Наиболее эффективный способ визуализации графов – это модель узлов-связей, где узлы представляют сущности, а связи представляют соединения.

Обычно узлы изображаются в виде маленьких точек или кругов. Связи отображаются в виде простых линий, соединенных между узлами. Однако в некоторых графах не все узлы и связи создаются одинаково: можно визуализировать дополнительные переменные, например, задав размер узла или интенсивность связи пропорциональными заданному значению.

Визуализация графа может быть использована для интерпретации структуры сети путем поиска любой кластеризации узлов, того, насколько плотно узлы связаны или как устроена система связей.

Двумя известными типами визуализации графа являются «ненаправленные» и «направленные». Ненаправленные графы отображают только соединения между объектами, в то время как направленные показывают с помощью маленьких стрелок, являются ли соединения односторонними или двусторонними.

Графы имеют ограниченный объем данных, потому становятся трудночитаемыми, когда узлов слишком много, и напоминают густо сплетенную паутину.

Для конечных пользователей визуализация графов является интуитивно понятной. Данный тип визуализации показывает, как сущности взаимосвязаны с помощью узлов и линий связи, чтобы представить их связи и помочь раскрыть тип отношений между группой объектов. Выход за рамки табличных и агрегированных представлений расширяет спектр возможностей анализа, который в противном случае был бы трудоемким.

Таким образом, можно выделить четыре основных преимущества графа как метода визуализации данных:

– простота – модель «узел-связь» сразу становится понятной даже для людей, которые никогда раньше не работали с графами;
– быстрота – люди отлично замечают закономерности, когда данные представлены в осязаемом формате. Выявление тенденций происходит быстрее при помощи визуализации графов;
– масштабируемость – графическая визуализация позволяет аналитикам видеть не только отдельные узлы данных, но и их контекст, структуру и отдельные связи;
– интерактивность – визуализации, доступные к взаимодействию, позволяют аналитикам, посредством обращения к данным, исследовать связанные объекты, раскрывая ранее скрытую информацию.

## *Сравнение подходов к визуализации графа*

Основным форматом визуализации графов является двухмерный (2D). Использование отрисовки данного типа позволяет быстро визуализировать и проанализировать небольшой объем данных, при этом не предъявляя особых технических требований к компьютеру (рис. 1). Однако при отображении большого количества узлов и связей возникает эффект «наложения».

В связи с расширением технических возможностей вычислительных машин особый рывок получает использование 3D-отрисовки [Oh-Hyun Kwon, Tarik Crnovrsanin, Kwan-Liu Ma 2017].

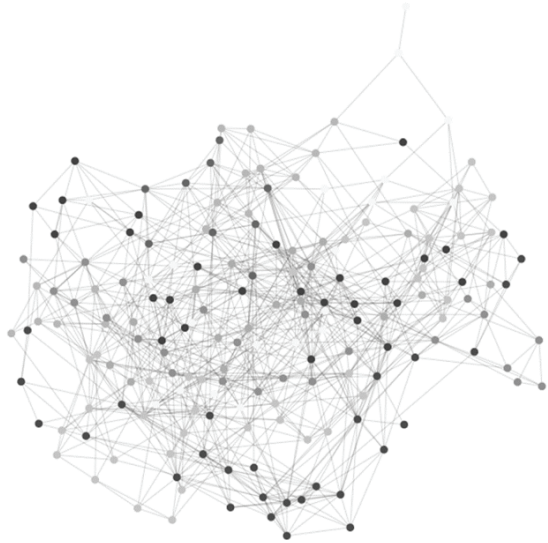

*Рис. 1.* Визуализация графа в формате 2D

Дополнительное третье измерение позволяет более гибко размещать элементы графа и вообще обходиться без пересечений ребер, но приводит к новым проблемам [Касьянов, Касьянова 2014]. На современном этапе невозможно обойтись без переходного состояния, поскольку современные устройства вывода по своей природе являются двумерными и поддерживают ограниченные разрешение и область визуализации, в связи с чем получается плоская проекция 3D-графа (рис. 2). Именно поэтому применение данного способа накладывает дополнительные требования к компьютерной системе.

В нашем исследовании были использованы следующие средства визуализации 2D и 3D: Force-graph & 3D Force-Directed Graph. Преимущество этого решения заключается в том, что отрисовка графов осуществляется не на основании мощностей центрального процессора компьютера, а с применением технологии WebGL, а значит, не нагружает видеокарту. На мощных компьютерах это позволяет отрисовывать графы до нескольких тысяч узлов.

*Рис. 2.* Визуализация графа в формате 3D

Таким образом, сравнение подходов к визуализации графа показало, что при отрисовке большого количества узлов (30 и более) изображение становится менее наглядным и 3D-формат более удобен. Однако данный подход невозможно назвать универсальным ввиду большого количества прочих потребностей аналитика при планировании (предметная область, формат и тип) и реализации (временные ресурсы, технические характеристики компьютера, библиотеки) исследования.

## *Математические методы представления и анализа графа*

В рамках одного из этапов разработки методики применения визуализации графов были собраны данные неформальных социальных (дружеских) связей студентов вуза в социальной сети.

Студенческое сообщество выступило объектом анализа не случайно. Современное поколение наиболее активно проявляет себя в сети. Молодежь генерирует колоссальное количество цифровой информации (контента), которая преимущественно локализуется в социальных сетях и мессенджерах, поскольку средства интернет-коммуникаций являются наиболее простым способом трансляции

авторского контента. Существенная часть ежедневной коммуникации молодежи осуществляется в формате онлайн. Дружеские связи молодых людей не только отражаются в социальных сетях, но нередко порождаются ими. Можно предположить, что профили пользователей социальных сетей младше 30 лет, в частности студентов, представляют собой достоверное отражение их взаимоотношений в реальной жизни.

В контексте сетевого взаимодействия отметим, что особую роль имеет не только общение между пользователями, но и возможность просматривать информацию друг о друге на стене профиля. Поэтому после сбора данных дружеских связей встала задача их агрегации до уровня групп для оценки интенсивности информационного потока.

Первоначальный алгоритм агрегации состоял в установлении веса связи между двумя группами, равного количеству дружеских связей между студентами данных групп:

$$F_{abs}(g_1, g_2) = m(g_1, g_2), \quad (1)$$

где $F_{abs}(g_1, g_2)$ – интенсивность информационного потока между двумя группами в абсолютном выражении;

$m(g_i, g_j)$ – количество дружеских связей между профилями студентов групп $g_i$ и $g_j$.

Такой алгоритм реализует метод абсолютной суммы связей. Принцип работы алгоритма приведен на изображении (рис. 3).

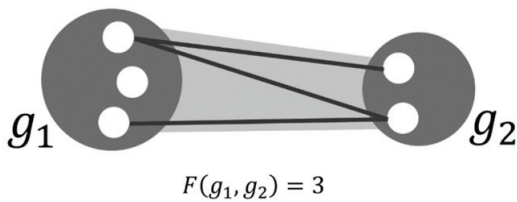


*Рис. 3.* Агрегация методом абсолютной суммы

Данный метод позволяет построить граф на основании максимально возможного абсолютного количества информации, передаваемого между студентами групп. Однако визуализация данных, агрегированных таким методом, недостаточно эффективна, если приходится сравнивать интенсивность информационных потоков между группами различных размеров [Проноза, Виткова, Чечулин, Котенко, Сахаров 2018]. Малые группы теряются на фоне больших.

Для разрешения этой проблемы был использован альтернативный метод агрегации связей между студентами групп. Он отличается от первого нормировкой выходного значения в диапазоне [0; 100] процентов. В данном случае учитывается заполненность множества потенциально возможных связей между студентами двух групп и используется эвристика квадратного корня для смягчения эффекта комбинаторного роста мощности множества потенциальных связей с ростом числа студентов двух групп. Метод описывается формулой

$$F(g_1, g_2) = \sqrt{\frac{m(g_1, g_2)}{n_{g1} * n_{g2}}} \cdot 100\%, \quad (2)$$

где $F(g_1, g_2)$ – интенсивность информационного потока между двумя группами;

$n_{gi}$ – количество студентов группы $g_i$;

$m(g_1, g_2)$ – количество дружеских связей между профилями студентов групп $g_i$ и $g_j$.

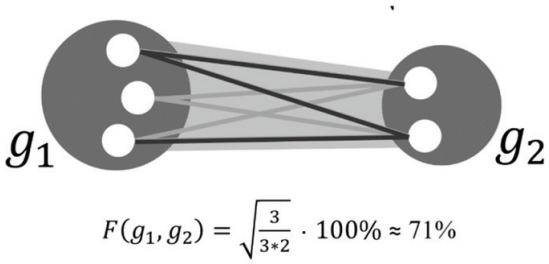


*Рис. 4.* Агрегация методом нормированной суммы

По итогам проведенной агрегации отметим, что анализ социальных связей в студенческой среде корректен при допущении, что обучение происходит в очном формате. Это следует из необходимости минимального личного взаимодействия для развития информационного потока.

Для проведения углубленного анализа авторами было введено понятие «рейтинг внешней социализации» ($R_i$) и разработана математическая модель по вычислению его уровня. Оценка данного показателя имеет особое значение при исследовании коммуникаций в молодежной среде, поскольку, как было отмечено ранее, сетевое взаимодействие все больше отражает реальное.

При подсчете метрики используется пропорциональное соотношение количества студентов в конкретной группе и количества внешних связей с другими группами:

$$R_i = \frac{m(g_i, g_n)}{n_{gi}}, \quad (3)$$

где $R_i$ – рейтинг внешней социализации конкретной группы;
$n_{gi}$ – количество студентов группы $g_i$;
$m(g_i, g_n)$ – количество дружеских связей между профилями студентов групп $g_i$ и $g_n$.

Представленная метрика позволяет осуществлять оценку внешней социализации студенческой группы сразу на нескольких уровнях: кафедры, факультета и всего университета.

Таким образом, математические методы представления и анализа графа позволяют скорректировать погрешность собранных данных, тем самым снижая расхождение между идеальным и реальным представлениями о коммуникационной структуре студенчества.

## *Возможности применения методики*

Студенты являются наиболее активными пользователями социальных сетей, поскольку имеют потребности двух типов. Они заключаются не только в получении знаний, но и в обмене мнениями и переживаниями друг с другом [Ермаков, Журавлев, Ковалевская 2016]. Такая гипотеза получает свое подтверждение в результате анализа собранных данных о дружеских связях студентов.

Проведенная авторами оценка интенсивности дружеских связей на разных уровнях может быть использована для понимания того, студенты каких направлений подготовки имеют наибольшее число дружеских связей. Это поможет распознать потоки данных, протекающих в студенческой среде.

Примером замкнутого графа с наибольшей интенсивностью связей может послужить визуализация данных факультета, на котором обучаются студенты-медтехники (рис. 5). Здесь две кафедры имеют крайне тесные социальные связи с первого курса, что может говорить о том, что все занятия там проходят совместно. Формат проведения лабораторных работ в силу специальности предполагает организацию коллективного взаимодействия. Также можно предположить, что в силу узкой направленности факультета студенты попадают в сообщество и знакомятся еще до фактического поступления в университет.

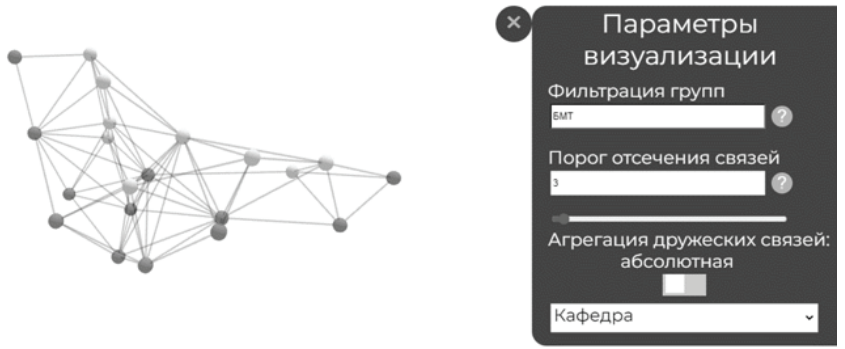


*Рис. 5.* Замкнутый граф дружеских связей студентов-медтехников

К неожиданным результатам привел анализ связей групп студентов-фундаменталистов. Граф получился несвязный с очень большим количеством висячих вершин. Гипотетически, это может быть связано со спецификой учебных направлений факультета. Как правило, люди, склонные к изучению физики и математики, более замкнуты в себе, что находит отражение в данной визуализации.

Таким образом, направление подготовки действительно играет ключевую роль при рассмотрении коммуникационной модели. Результаты анализа структуры социальных связей студенческого сообщества могут быть полезны при формировании информационной и социальной политики вуза в отношении студентов и абитуриентов.

Как стало понятно, социальные сети представляют собой колоссальный по объему источник открытых данных о неформальных связях формальных социальных групп. Естественным образом возникает вопрос о возможных вариантах использования этих данных в интересах социальной аналитики.

Действительно, с распространением и развитием технологий обработки больших данных спрос на аналитику только растет. Это связано с задачами интерпретации результатов исследований. Интерпретация численных результатов обработки массивов данных – нетривиальная задача. В условиях работы с большими объемами неструктурированных данных исследователь зачастую не знает, что именно он ищет.

Для этого нужна интерактивная визуализация данных. Человек воспринимает через зрительный канал, по некоторым оценкам, до

90% информации [Алешин 2008]. Основываясь на этом, хочется подчеркнуть, что визуализация графа является одним из самых эффективных способов представления данных.

## *Заключение*

Данные социальных сетей разнообразны. Они содержат информацию о структурах дружеских связей между людьми, их интересах. Публикуемые пользователями материалы отражают их личный опыт и социально-политическую позицию, представляя большой интерес для социальных исследователей [Ремарчук 2022]. Социальные сети порождают принципиально новые подходы к изучению социума – неопросные и косвенные методы, потенциал которых еще только предстоит открыть.

Авторами были выделены четыре основных преимущества графа как метода визуализации данных, произведен сравнительный анализ подходов к визуализации графа, сделаны практические выводы по их применению. Математические методы представления и анализа графа позволили скорректировать погрешность собранных данных.

В качестве практического эксперимента, который позволил показать возможности приложения средств визуализации графа к анализу данных социальных сетей, была выбрана интерактивная визуализация социальной структуры студенческого сообщества вуза на основании данных социальной сети.

Полученная методика применения визуализации графа для анализа неформальных социальных (дружеских) связей может быть использована во многих сферах общества. Задачи визуализации и анализа в ближайшее время и в долгосрочной перспективе будут иметь высокий спрос и немалый объем инвестиций.

## *Литература*

Алешин 2008 – *Алешин Л.И.* Информационные технологии. М.: Литера, 2008. 424 с. (Современная библиотека; Вып. 35).

Долинина, Печенкин, Тарасова 2009 – *Долинина О.Н., Печенкин В.В., Тарасова В.В.* Использование графовых моделей для визуализации социальных сетей образовательной организации // Вестник СГТУ. 2009. Т. 4. № 2 (43). С. 210–214.

Ермаков, Журавлев, Ковалевская 2016 – *Ермаков В.А., Журавлев Г.Т., Ковалевская Е.В.* Анализ активности студентов в социальных сетях Интернета // Интерактивная наука. 2016. № 8. С. 44–48.

Касьянов, Касьянова 2014 – *Касьянов В.Н., Касьянова Е.В.* Визуализация информации на основе графовых моделей: Учеб. пособие. Новосибирск: НГУ, 2014. 148 с.

Линник 2018 – *Линник Е.В.* Графовая аналитика для решения ключевых проблем в банковской сфере // Молодой ученый. 2018. № 52 (238). С. 128–134.

Мельникова, Яковлев 2014 – *Мельникова М.С., Яковлев И.П.* Понятие «Социальная сеть» в социологических теориях и интернет-практиках // Вестник СПбГУ. Серия 9. Филология. Востоковедение. Журналистика. 2014. № 1. С. 254–257.

Проноза, Виткова, Чечулин, Котенко, Сахаров 2018 – *Проноза А.А., Виткова Л.А., Чечулин А.А., Котенко И.В., Сахаров Д.В.* Методика выявления каналов распространения информации в социальных сетях // Вестник Санкт-Петербургского ун-та. Прикладная математика. Информатика. Процессы управления. 2018. Т. 14. Вып. 4. С. 362–377.

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

Соловьев, Гончаров, Батин 2020 – *Соловьев С.А., Гончаров С.С., Батин Р.Е.* Методика определения структуры неформальных информационных потоков в студенческой среде по данным социальных сетей на примере МГТУ им. Н.Э. Баумана // Аналитические технологии в социальной сфере: теория и практика: Сб. статей VIII студенческой научной конференции, Москва, 27 мая 2020 года. М.: Научно-исследовательский центр проблем национальной безопасности, 2020. С. 29–33.

Oh-Hyun Kwon, Tarik Crnovrsanin, Kwan-Liu Ma 2017 – *Oh-Hyun Kwon, Tarik Crnovrsanin, Kwan-Liu Ma*. What Would a Graph Look Like in This Layout? A Machine Learning Approach to Large Graph Visualization // IEEE Transactions on Visualization and Computer Graphics, 11 oct. 2017. Piscataway, NJ: IEEE Standards Office, 2017.

## *References*

Aleshin, L.I. (2008), *Informatsionnye tekhnologii* [Information Technologies], Litera, Moscow, Russia, 424 p. (Modern Library; Issue 35).

Dolinina, O.N., Pechenkin, V.V. and Tarasova, V.V. (2009), “The use of graph models for visualization of social networks in educational organizations”, Vestnik SSTU, vol. 4, no. 2 (43), pp. 210–214.

Ermakov, V.A., Zhuravlev, G.T. and Kovalevskaya, E.V. (2016), “Analysis of student activity in Internet social networks”, *Interactive science*, no. 8, pp. 44–48.

Kas’yanov, V.N. and Kas’yanova, E.V. (2014), *Vizualizatsiya informatsii na osnove grafovykh modelei: ucheb. posobie* [Visualization of information based on graph models. Study guide], NSU, Novosibirsk, Russia, 148 p.

Linnik, E.V. (2018), “Graph analytics for solving key issues in the banking sector”, *Young scientist*, vol. 52 (238), pp. 128–134.

Mel'nikova, M.S. and Yakovlev, I.P. (2014), "The concept of "Social network" in sociological theories and Internet practices", *Vestnik SPbSU. Series 9. Philology. Oriental Studies. Journalism*, vol. 1, pp. 254–257.

Pronoza, A.A., Vitkova, L.A., Chechulin, A.A., Kotenko, I.V. and Sakharov, D.V. (2018), "Methodology for identifying information distribution channels in social networks", *Vestnik SPbSU. Applied Mathematics. Computer Science. Control Processes*, vol. 14, issue 4, pp. 362–377.

Remarchuk, V.N. (2022), *Informatsionnaya analitika: teoriya, metodologiya, tekhnologii: uchebnik* [Information analytics. Theory, methodology, technologies. Textbook], Lan, St. Petersburg, Russia, 224 p.

Solov'ev, S.A., Goncharov, S.S. and Batin, R.E. (2020), "Methodology for determining the structure of informal information flows in the student surroundings based on social network data on the example of Bauman Moscow State Technical University" *Analiticheskie tekhnologii v sotsial'noi sfere: teoriya i praktika: Sb. statei VIII studencheskoi nauchnoi konferentsii, Moskva, 27 maya 2020 goda* [Analytical technologies in the social sphere. Theory and practice. Collection of proceedings of the 8th Student Scientific Conference, Moscow, May 27, 2020, Research Center for National Security Issues, Moscow, Russia, pp. 29–33.

Oh-Hyun Kwon, Tarik Crnovrsanin and Kwan-Liu Ma (2017), "What Would a Graph Look Like in This Layout? A Machine Learning Approach to Large Graph Visualization", IEEE Transactions on Visualization and Computer Graphics, 11 October 2017, IEEE Standards Office, Piscataway, NJ, USA.

## *Информация об авторах*

*Дмитрий Г. Алпацкий*, кандидат политических наук, Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; alpackij@bmstu.ru

*Алексей А. Канаев*, студент, Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; alekseykanaev@mail.ru

*Андрей И. Колбин*, студент, Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; kolbinai@student.bmstu.ru

*Дмитрий Д. Ставский*, студент, Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия; 105005, Россия, Москва, 2-я Бауманская ул., д. 5; stavskiydd@student.bmstu.ru

*Information about the authors*

*Dmitrii G. Alpatskii*, Cand. of Sci. (Political Sciences), Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, Russia, 105005; alpackij@bmstu.ru

*Aleksei A. Kanaev*, student, Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, Russia, 105005; alekseykanaev@mail.ru

*Andrei I. Kolbin*, student, Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, Russia, 105005; kolbinai@student.bmstu.ru

*Dmitrii D. Stavskii*, student, Bauman Moscow State Technical University, Moscow, Russia; bld. 5, 2nd Bauman Str., Moscow, Russia, 105005; stavskiydd@student.bmstu.ru

УДК 004.81
DOI: 10.28995/2686-679X-2023-1-23-37

# Многофакторный когнитивный анализ защищенности объектов культурно-исторического наследия России

Евгений Н. Надеждин
*Российский государственный гуманитарный университет, Москва, Россия, en-hope@yandex.ru*

*Аннотация*. В данной статье изучается проблема сохранения культурно-исторического наследия городов России в условиях существенного расширения спектра угроз, изменения условий финансирования и хозяйственной деятельности. Анализ и ранжирование факторов риска нарушения внешнего облика и характеристик объектов культурно-исторического наследия позволяют определить наиболее эффективные методы и средства ослабления угроз и разработать оптимальную стратегию защиты. Применение аналитических методов факторного анализа для решения указанных задач затруднительно в силу качественного характера и разнородности факторов риска. Показано, что для идентификации предметной области может быть применена методика когнитивного моделирования и многофакторного анализа на основе нечеткой когнитивной карты, которая представляет собой семантическую сеть специального вида. Для иллюстрации предлагаемого подхода в качестве объекта исследования выбран Государственный мемориальный и природный заповедник музей-усадьба Л.Н. Толстого «Ясная Поляна», расположенный в Тульской области. Для численного решения задачи использована авторская методика многофакторного когнитивного анализа, основанная на применении нечетких когнитивных карт В.Б. Силова. При идентификации нечеткой когнитивной карты выделены три группы концептов: факторы угроз; методы и средства защиты; целевые факторы. После преобразования исходной когнитивной матрицы к транзитивно-замкнутой форме вычислены системные показатели нечеткой когнитивной карты. В итоге определены концепты, оказывающие доминирующее весовое влияние на защищенность музея-усадьбы «Ясная Поляна». Результаты исследования позволяют обосновать комплекс эффективных мероприятий, направленных на сохранение избранного объекта культурно-исторического наследия.

*Ключевые слова:* культурно-историческое наследие, проблема сохранения культурно-исторического наследия, защищенность, многофакторный анализ, нечеткая когнитивная карта

*Для цитирования: Надеждин Е.Н.* Многофакторный когнитивный анализ защищенности объектов культурно-исторического наследия России // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 1. С. 23–37. DOI: 10.28995/2686-679X-2023-1-23-37

# Multivariate cognitive analysis of protecting the objects of cultural and historical heritage in Russia

Evgenii N. Nadezhdin
*Russian State University for the Humanities, Moscow, Russia, en-hope@yandex.ru*

*Abstract*. This article studies the issue of preserving the cultural and historical heritage of Russian cities in the face of a significant expansion of the range of threats, changes in the conditions of financing and economic activity. Analysis and ranking of risk factors for violation of the appearance and characteristics of objects of cultural and historical heritage make it possible to determine the most effective methods and means of mitigating threats and develop an optimal protection strategy. The application of analytical methods of factor analysis to solve those issues is difficult due to the qualitative nature and heterogeneity of risk factors. It is shown that the method of cognitive modeling and multivariate analysis based on a fuzzy cognitive map, which is a special type of semantic network, can be used to identify the subject area. To illustrate the proposed approach, the author opted for the State Memorial and Natural Reserve Museum-estate of L.N. Tolstoy "Yasnaya Polyana", located in the Tula region. For the numerical solution of the issue, the author's method of multivariate cognitive analysis based on the use of fuzzy cognitive maps by V.B. Silov. When identifying a fuzzy cognitive map, three groups of concepts were identified: threat factors; methods and means of protection; target factors. After transforming the original cognitive matrix to a transitive-closed form, the system indicators of the fuzzy cognitive map are calculated. As a result, the concepts that have a dominant weight influence on the protection of the Museum-estate "Yasnaya Polyana" are determined. The results of the study make it possible to substantiate a set of effective measures aimed at preserving the chosen object of cultural and historical heritage.

*Keywords:* cultural and historical heritage of Russia, the issue of preserving objects of cultural and historical heritage, the issue of factor analysis, fuzzy cognitive maps

*For citation:* Nadezhdin, E.N. (2023), "Multivariate cognitive analysis of protecting the objects of cultural and historical heritage in Russia", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 1, pp. 23–37, DOI: 10.28995/2686-679X-2023-1-23-37

Как известно, культурно-историческое наследие народа во многом определяет его менталитет, преемственность гуманитарных ценностей и сохраняет национальные традиции [Драчева 2016; Шаманская 2014]. Объекты культурного наследия представляют собой уникальную ценность для многонационального народа России и являются неотъемлемой частью всемирного культурного наследия. Сохранение культурно-исторического наследия – основа дальнейшего развития общества, это конституционная обязанность каждого гражданина страны. Однако, по данным статистики, физическое состояние многих находящихся под охраной государства памятников истории и культуры России в наше время оценивается как неудовлетворительное [Сиволап 2012]. Эта ситуация требует разработки научного подхода к изучению угроз и формированию комплекса первоочередных мер по защите объектов культурно-исторического наследия [Исаев, Гуркина 2013].

В настоящее время не существует универсальных методик комплексной оценки и прогнозирования показателей защищенности объектов культурно-исторического наследия (КИН). В условиях мирового экономического кризиса и стагнации промышленного производства для факторного анализа защищенности объектов КИН является особенно актуальной.

*Целью статьи* является обоснование методического подхода к проблеме сохранения объектов культурно-исторического наследия России и его апробация на примере когнитивного моделирования и анализа защищенности музея-усадьбы Л.Н. Толстого «Ясная Поляна» с применением нечетких когнитивных карт.

В нашем исследовании в качестве объекта КИН рассматривается Государственный музей-усадьба Л.Н. Толстого «Ясная Поляна», расположенный в Щёкинском районе Тульской области [Кулешова 2020]. Сегодня усадьба Л.Н. Толстого является музеем, государственным мемориальным и природным заповедником, входящим в список объектов культурного наследия РФ регионального значения. На рис. 1–4 показаны фотографии фрагментов музея-усадьбы Л.Н. Толстого «Ясная Поляна».

Для факторного анализа проблемы сохранения избранного объекта КИН России воспользуемся известной методикой когнитивного анализа слабоструктурированных систем на основе аппарата нечетких когнитивных карт (НКК) В.Б. Силова [Силов 1995] и рекомендаций авторских работ [Nadezhdin 2016; Шершакова, Надеждин 2018]. Указанный вид НКК удачно сочетает достоинства экспертных оценок и метода идентификации на основе когнитивных карт и обеспечивает корректность формального описания слабоструктурированной предметной области и приемлемую для практики точность идентификации.

*Рис. 1.* Имение Л.Н. Толстого в дер. Ясная Поляна

*Рис. 2.* Имение Л.Н. Толстого в дер. Ясная Поляна

*Рис. 3.* Обеденный зал в Ясной Поляне

*Рис. 4.* Рабочий кабинет писателя Л.Н. Толстого

Понятие нечеткой когнитивной карты В.Б. Силова представляет собой расширение классического понятия когнитивной карты, основанное на предположении, что взаимовлияния между концептами могут различаться по интенсивности и их интенсивность может изменяться с течением времени. Для этого в НКК вводят показатель интенсивности влияния и от классического отношения переходят к нечеткому отношению $W$, элементы $w_{i,j}$ которого характеризуют направление и степень интенсивности (вес) влияния между концептами $e_i$ и $e_j$:

$$w_{i,j} = w(e_i, e_j),$$

где $w$ – нормированный показатель интенсивности влияния (характеристическая функция отношения $W$), обладающий рядом специальных свойств [Силов 1995]. Среди этих свойств выделим следующее: $w_{i,j}$ принимает значение из интервала $[-1; 1]$ при промежуточной степени положительного или отрицательного влияния.

Нечеткая когнитивная карта отображает исследуемый объект в виде взвешенного ориентированного графа, вершины которого соответствуют элементам множества $E$ (концептам), а дуги – ненулевым элементам отношения $W$, т. е. причинно-следственным связям. Каждая дуга имеет вес, задаваемый соответствующим значением $w_{i,j}$. Отношение $W$ представимо в виде когнитивной матрицы $W = \{w_{i,j}, i, j = \overline{1, n}\}$ размерности $(n \times n)$ ($n$ – число концептов в системе), которая будет интерпретироваться как матрица смежности данного графа. *Состояние системы* в текущий момент времени определяется набором значений всех концептов НКК. Целевое состояние системы задается вектором значений множества целевых концептов.

Предположим, что априорно экспертами определен набор базовых факторов (концептов), оказывающих существенное влияние на состояние защищенности принятого объекта исследования (табл. 1). Укажем характерные этапы когнитивного анализа механизма обеспечения защищенности КИН.

Этап 1. *Построение НКК механизма защиты объекта КИН*. Для этого используют эвристические данные, полученные при реализации специальных процедур извлечения и обработки экспертной информации.

На *первом шаге* в результате опроса шести экспертов выделено 12 существенных факторов (далее – *концептов*), определяющих защищенность КИН. Эти концепты условно разделены на три группы (см. табл. 1): а) угрозы и деструктивные воздействия; б) методы и средства защиты; в) целевые факторы.

На *втором шаге* эксперты указали причинно-следственные связи между концептами с выделением их характера (направленности). Результатом таких действий стало построение когнитивной карты (см. рис. 5), формально отражающей причинно-следственные связи без учета интенсивности взаимовлияний концептов.

На *третьем шаге* эксперты привлекались для оценки силы влияния концептов друг на друга. В итоге по результатам исследований была построена когнитивная матрица $W = \{w_{i,j}, i, j = \overline{1, 12}\}$ (табл. 2), дополняющая когнитивную карту (см. рис. 5). Элементы когнитивной матрицы определялись как усредненные по числу экспертов оценки интенсивностей влияния концептов друг на друга. В полученной НКК представлены наиболее важные, непосредственные связи между концептами.

Для когнитивного анализа причинно-следственной структуры и характеристик механизма защиты КИН необходима информация о неявных проявлениях влияния концептов друг на друга и на конечный результат. Примером такого влияния может служить цепочка концептов, характеризуемая фрагментом пути в когнитивном графе (см. рис. 5): $e_4 \rightarrow e_6 \rightarrow e_7 \rightarrow e_{12}$.

Этап 2. *Количественная оценка опосредованного взаимовлияния концептов*. Для этого на первом шаге выполняют операцию транзитивного замыкания когнитивной матрицы. Из множества известных способов транзитивного замыкания матрицы смежности воспользуемся алгоритмом, рекомендуемым в работе [Силов 1995, с. 99]. Алгоритм заключается в следующем.

*Таблица 1*

Сводная матрица концептов когнитивной модели

| № пп | Наименование концепта | Идентификатор концепта |
|---|---|---|
| А. Угрозы и деструктивные воздействия | | |
| 1 | Экономический кризис и изменение приоритетов в духовной жизни гражданского общества | $e_1$ |
| 2 | Интенсивное развитие промышленности и бизнеса. Коммерческая застройка охранных зон | $e_2$ |
| 3 | Конкуренция в области туристического бизнеса | $e_3$ |
| 4 | Влияние времени и экологических факторов на физическое состояние | $e_4$ |
| 5 | Туристический вандализм | $e_5$ |
| Б. Методы и средства защиты | | |
| 6 | Научные исследования. Мониторинг состояния объекта. Анализ и учет общественного мнения | $e_6$ |
| 7 | Восстановление, реставрация и реконструкция объектов с применением современных технологий | $e_7$ |
| 8 | Маркетинговая стратегия | $e_8$ |
| 9 | Поддержка государства и общественных организаций. Комплекс правовых мер защиты | $e_9$ |
| 10 | Профессиональная компетентность, научная квалификация и мотивированность штатных сотрудников | $e_{10}$ |
| В. Целевые концепты. Ожидаемые результаты | | |
| 11 | Интеграция объекта в систему культурного туристического бизнеса | $e_{11}$ |
| 12 | Физическое состояние и привлекательность объекта. Уровень защищенности объекта | $e_{12}$ |

*Таблица 2*

Когнитивная матрица $W = \{w_{ij}, i, j = \overline{1, 12}\}$ механизма защиты

| Номер концепта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0,20 | 0 | 0 | 0,30 | 0 | 0 | –0,45 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0,30 | 0 | 0 | –0,45 | 0 | 0 | 0 | –0,32 | –0,25 |
| 3 | 0 | 0 | 0 | 0 | 0,10 | 0 | –0,10 | 0 | 0 | 0 | –0,15 | 0 |
| 4 | 0 | 0 | 0,25 | 0 | 0 | 0,26 | –0,35 | –0,10 | 0 | 0 | –0,15 | –0,10 |
| 5 | 0 | 0 | 0 | 0,10 | 0 | 0 | –0,40 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0,55 | 0,45 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,36 | 0 | 0 | 0,55 | 0,30 |
| 8 | 0 | 0 | –0,30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,60 | 0 |
| 9 | –0,20 | –0,55 | 0 | 0 | –0,45 | 0,20 | 0,60 | 0 | 0 | 0,25 | 0,65 | 0,35 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,75 | 0,85 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,60 |
| 12 | 0 | 0 | 0 | 0 | –0,40 | 0 | 0 | 0,20 | 0 | 0 | 0 | 0 |

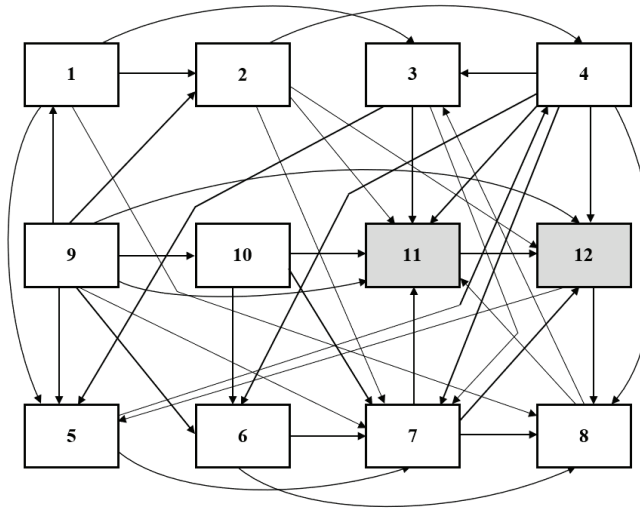


*Рис. 5*. Когнитивная карта механизма обеспечения защищенности объекта культурно-исторического наследия – музея-заповедника «Ясная Поляна»

1. От исходной когнитивной матрицы (табл. 3) переходят к когнитивной матрице положительных связей $R$ размерностью $(2 \cdot n \times 2 \cdot n)$ на основе процедуры замены:

$$\text{если } w_{i,j} > 0, \text{ то } r_{2i-1,\,2j-1} = w_{i,j},\ r_{2i,\,2j} = w_{i,j};$$

$$\text{если } -w_{i,j} < 0, \text{ то } r_{2i-1,\,2j} = -w_{i,j},\ r_{2i,\,2j-1} = -w_{i,j}.$$

Остальные элементы матрицы $R$ принимают нулевое значение.

2. *Определяют транзитивное замыкание нечеткого отношения* $R$ в соответствии с выражением [Silov 1995, p. 29]

$$\widetilde{R} = \bigcup_{i=1}^{n} R^i = R \cup R^2 \cup \ldots \cup R^n, \text{ где } R^2 = R \times R.$$

Произведение нечетких отношений вычисляют согласно процедуре:

$$\text{если } D = A + B, \text{ то } d_{i,j} = \max_{k=1,\ldots n} a_{i,k} \cdot b_{k,j},\ i,j = 1,\ldots n.$$

3. От вспомогательной матрицы $\widetilde{R}$ переходят к транзитивно замкнутой когнитивной матрице $V$, элементами которой будут пары $(v_{i,j}, \widetilde{v}_{i,j})$, где $v_{i,j}$ и $\widetilde{v}_{i,j}$ характеризуют соответственно силу положительного и отрицательного влияния $i$-го концепта на $j$-й концепт:

$$v_{i,j} = \max(r_{2i-1,\,2j-1}, r_{2i,\,2j}); \quad \widetilde{v}_{i,j} = -\max(r_{2i-1,\,2j}, r_{2i,\,2j-1}).$$

4. Вычисление двух основных показателей НКК:

а) воздействие $i$-го концепта на $j$-й концепт:

$$h_{i,j} = \operatorname{sgn}(v_{i,j} + \widetilde{v}_{i,j}) \cdot \max(|v_{i,j}|, |\widetilde{v}_{i,j}|) \quad |v_{i,j}| \neq |\widetilde{v}_{i,j}|; \tag{1}$$

б) консонанс влияния $i$-го концепта на $j$-й концепт:

$$c_{i,j} = \frac{|v_{i,j} + \widetilde{v}_{i,j}|}{|v_{i,j}| + |\widetilde{v}_{i,j}|}, \tag{2}$$

который выражает меру доверия к знаку воздействия.

С использованием полученных данных определяют интегральные показатели консонанса и воздействия [Силов 1995, с. 102].

Этап 3. *Анализ результатов расчетов и обоснование рекомендаций.*

В соответствии с приведенными выше расчетными соотношениями на основании полученной транзитивно замкнутой когнитивной матрицы $V = [(n_{i,j}, \tilde{n}_{i,j}), i, j = \overline{1, n}]$ вычисляют частные (см. выражения (1) и (2)) и системные показатели нечеткой когнитивной модели.

В табл. 3 представлены значения системных показателей нечеткой когнитивной модели механизма защиты объекта КИН (далее – системы).

Результаты расчетов свидетельствуют о доминирующем положительном влиянии концептов $e_6$, $e_7$ и $e_{10}$ и отрицательном влиянии концептов $e_1$, $e_2$ и $e_5$ на механизм защиты КИН. Дадим краткие пояснения к результатам когнитивного моделирования.

*Концепт 1 «Экономический кризис и изменение приоритетов в духовной жизни гражданского общества»* оказывает наиболее сильное отрицательное влияние на систему ($\vec{P}_1 = -0{,}044$), а система, в свою очередь, имеет незначительное влияние на концепт ($\overleftarrow{P}_1 = -0{,}017$). При этом средний по значению консонанс концепта

($\vec{H}_1 = 0{,}416$) указывает на нестабильность отмеченной закономерности. По мере завершения активной фазы пандемии и выхода из затяжного экономического кризиса следует ожидать повышение интереса социально активной части гражданского населения России к изучению объектов национального КИН. В этом случае влияние концепта 1 на систему будет ослаблено.

*Концепт 2 «Интенсивное развитие промышленности и бизнеса. Коммерческая застройка охранных зон»* оказывает сильное влияние на систему ($\vec{P}_2 = -0{,}052$), а система, в свою очередь, стремится ослабить влияние концепта ($\overleftarrow{P}_2 = -0{,}046$). Данный эффект можно объяснить возможностью общественного контроля за процессом несанкционированного промышленного «освоения» охранных зон объектов КИН.

*Концепт 5 «Туристический вандализм»* ($\vec{P}_5 = -0{,}056$) является характерной чертой незначительной части молодого поколения, однако результаты его воздействия (например, лесные пожары) могут нанести серьезный ущерб объектам КИН. Для ослабления негативного влияния данного фактора на систему следует активизировать весь арсенал методов и средств общественного и правового воздействия на потенциальных нарушителей культурного ландшафта.

*Концепт 6 «Научные исследования. Мониторинг состояния объекта. Анализ и учет общественного мнения»* ($\vec{P}_6 = 0{,}106$) предполагает проведение системных научных исследований, направленных на изучение артефактов и истории объекта КИН в интересах популяризации народных традиций, определения физического состояния объекта и разработки перспективных методов и технологий его сохранения как национального достояния. Представляется, что данный концепт обладает большим созидательным потенциалом.

*Концепт 7 «Восстановление, реставрация и реконструкция объектов с применением современных технологий»* ($\vec{P}_7 = 0{,}086$) имеет существенное значение для сохранения первоначального облика объектов КИН. Отметим также, что обоснованное применение достижений в области новейших технологий (строительных, дизайнерских, осветительных, информационных и др.) позволяет укрепить древние конструкции, повысить привлекательность объектов КИН и интегрировать их в современный городской и природный ландшафт.

*Концепт 10 «Профессиональная компетентность, научная квалификация и мотивированность штатных сотрудников»* ($\vec{P}_{10} = 0{,}123$) следует отнести к числу наиболее значимых факторов,

которые определяют современный физический облик и коммерческий потенциал объекта КИН в контуре туристического бизнеса. Через активное общение с ведущими специалистами и экспертами посетители получают уникальные знания, которые создают эффект положительной обратной связи, имеющий важное значение для духовного развития личности и усиления национального самосознания.

*Таблица 3*

Результаты расчета системных показателей нечеткой когнитивной модели

| № п.п. | Показатели консонанса | | Показатели диссонанса | | Показатели влияния | | Показатель централизации влияния |
|---|---|---|---|---|---|---|---|
| | $\vec{H}_i$ | $\overleftarrow{H}_j$ | $\vec{D}_i$ | $\overleftarrow{D}_j$ | $\vec{P}_i$ | $\overleftarrow{P}_j$ | $E_i$ |
| 1 | 0,416 | 0,083 | 0,584 | 0,917 | –0,044 | –0,017 | –0,027 |
| 2 | 0,572 | 0,083 | 0,428 | 0,917 | –0,052 | –0,046 | $-6{,}6\cdot10^{-3}$ |
| 3 | 0,651 | 0,906 | 0,349 | 0,094 | –0,021 | –0,020 | $-9{,}8\cdot10^{-4}$ |
| 4 | 0,324 | 0,925 | 0,676 | 0,075 | –0,019 | 0,012 | –0,031 |
| 5 | 0,618 | 0,908 | 0,382 | 0,092 | –0,056 | –0,098 | 0,042 |
| 6 | 0,664 | 0,944 | 0,336 | 0,056 | 0,106 | 0,045 | 0,061 |
| 7 | 0,658 | 0,871 | 0,342 | 0,129 | 0,086 | 0,019 | 0,067 |
| 8 | 0,655 | 0,834 | 0,345 | 0,166 | 0,052 | 0,056 | $-3{,}5\cdot10^{-3}$ |
| 9 | 0,830 | 0 | 0,170 | 1 | 0,073 | 0 | 0,073 |
| 10 | 0,651 | 0,083 | 0,349 | 0,917 | 0,123 | 0,021 | 0,102 |
| 11 | 0,648 | 0,843 | 0,352 | 0,157 | 0,048 | 0,158 | –0,109 |
| 12 | 0,644 | 0,850 | 0,356 | 0,150 | 0,003 | 0,170 | –0,166 |

Опираясь на результаты многофакторного когнитивного анализа, выполненного по авторской методике [Nadezhdin 2016], можно предположить, что наибольший положительный эффект следует ожидать от согласованного изменения группы управляемых концептов НКК, которые находятся в цепочке причинно-следственной связи и в совокупности обеспечивают устойчивое положительное воздействие на систему обеспечения защищенности объекта КИН. Характерным примером здесь может служить цепочка концептов: $e_6 \rightarrow e_7 \rightarrow e_{12}$. Результаты критического анализа и взвешенной оценки потенциала указанного ресурса способны придать новый

импульс инновационным преобразованиям в системе управления защищенностью объектов культурно-исторического наследия.

## *Выводы*

1. В результате исследования показано, что в основу научно-обоснованного решения проблемы сохранности и защиты объектов КИН России может быть положена методика многофакторного когнитивного анализа с применением НКК.

2. В ходе исследования на примере музея-заповедника «Ясная Поляна» выявлены значимые негативные и позитивные факторы и количественно оценено их воздействие на сохранность объекта КИН. При наличии систематического мониторинга ситуации и согласованной настройки положительных концептов расширяются возможности для своевременной компенсации влияния негативных факторов.

3. Полученные результаты могут служить методической основой для дальнейших исследований, направленных на разработку комплекса первоочередных мер по сохранению объектов КИН регионального значения.

## *Литература*

Драчева 2016 – *Драчева Е.Л.* Проблемы сохранения объектов культурного наследия ЮНЕСКО в России и за рубежом. URL: http://futureruss.ru/wp-content/uploads/2016/05/ДРАЧЕВА_doc.pdf (дата обращения 20 октября 2022).

Исаев, Гуркина 2013 – *Исаев А.П., Гуркина Н.К.* Российская историография: проблемы сохранения культурного наследия // Управленческое консультирование. 2013. № 11 (59). URL: https://cyberleninka.ru/article/n/rossiyskaya-istoriografiya-problemy-sohraneniya-kulturnogo-naslediya (дата обращения 15 октября 2022).

Кулешова 2020 – *Кулешова М.Е.* Русская усадьба как перспективный проект всемирного наследия ЮНЕСКО: на примере «Ясной Поляны» Л.Н. Толстого // Наследие и современность. 2020. № 4. URL: https://cyberleninka.ru/article/n/russkaya-usadba-kak-perspektivnyy-obekt-vsemirnogo-naslediya-yunesko-na-primere-yasnoy-polyany-l-n-tolstogo (дата обращения 9 октября 2022).

Сиволап 2012 – *Сиволап Т.Е.* К вопросу сохранения культурного наследия в России: некоторые аспекты решения проблемы // Наука о человеке: гуманитарные исследования. 2012. № 1 (9). URL: https://cyberleninka.ru/article/n/k-voprosu-sohraneniya-kulturnogo-naslediya-v-rossii-nekotorye-aspekty-resheniya-problemy (дата обращения 23 октября 2021).

Силов 1995 – *Силов В.Б.* Принятие стратегических решений в нечеткой обстановке: Монография. М.: ИНПРО-РЕС, 1995. 228 с.

Шаманская 2014 – *Шаманская И.Ю.* Сохранение культурно-исторического наследия – условие устойчивого развития городов России // Актуальные проблемы гуманитарных и естественных наук. 2014. № 12–2. С. 298–300.

Шершакова, Надеждин 2018 – *Шершакова Т.Л., Надеждин Е.Н.* Когнитивный анализ защищенности информационных ресурсов образовательной организации // Информация и безопасность. 2018. Т. 21. Вып. 1. С. 48–57.

Nadezhdin 2016 – *Nadezhdin E.N.* Fuzzy cognitive model of the mechanism of support of competitiveness of the software product // Austrian Journal of Technical and Natural Sciences. Scientific journal. 2016. № 1–2 (January–February). P. 13–19.

## *References*

Dracheva, Ye.L. (2016), "Issues of preservation of UNESCO cultural heritage sites in Russia and abroad", available at: http://futureruss.ru/wp-content/uploads/2016/05/DRACHEVA_doc.pdf (Accessed 20 October 2022).

Isaev, A.P. and Gurkina, N.K. (2013), "Russian historiography. Issues of cultural heritage preservation", *Management consulting*, vol. 11 (59), available at: https://cyberleninka.ru/article/n/rossiyskaya-istoriografiya-problemy-sohraneniya-kulturnogo-naslediya (Accessed 15 October 2022).

Kuleshova, M.Ye. (2020), Russian estate as a promising UNESCO World Heritage project. The example of L.N. Tolstoy's "Yasnaya Polyana", *Heritage and the present time*, vol. 4, available at: https://cyberleninka.ru/article/n/russkaya-usadba-kak-perspektivnyy-obekt-vsemirnogo-naslediya-yunesko-na-primere-yasnoy-polyany-l-n-tolstogo (Accessed 9 October 2022).

Nadezhdin, E.N. (2016), "Fuzzy cognitive model of the mechanism of support of competitiveness of the software product", *Austrian Journal of Technical and Natural Sciences. Scientific journal*, vol. 1–2, (January–February), pp. 13–19.

Sivolap, T.Ye. (2012), "On preserving the cultural heritage in Russia. Some aspects of solving the issue", *Science of man. Humanitarian research*, vol. 1 (9), available at: https://cyberleninka.ru/article/n/k-voprosu-sohraneniya-kulturnogo-naslediya-v-rossii-nekotorye-aspekty-resheniya-problemy (Accessed 23 October 2022).

Silov, V.B. (1995), *Prinyatie strategicheskikh reshenii v nechetkoi obstanovke: monografiya* [Making strategic decisions in a fuzzy situation. Monograph], INPRO-RES, Moscow, Russia, p. 228.

Shamanskaya, I.Yu. (2014), "Retaining the cultural and historical heritage is a condition for the sustainable development of Russian cities", *Current issues of the humanities and natural sciences*, vol. 12–2, pp. 298–300.

Shershakova, T.L. and Nadezhdin, Ye.N. (2018), "Cognitive analysis of the information resources security in an educational organization", *Information and security*, vol. 21, issue 1, pp. 48–57.

*Информация об авторе*

*Евгений Н. Надеждин*, доктор технических наук, профессор, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; en-hope@yandex.ru

*Information about the author*

*Evgenii N. Nadezhdin*, Dr. of Sci. (Computer Science), professor, Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Sq., Moscow, Russia, 125047; en-hope@yandex.ru.

УДК 004.056
DOI: 10.28995/2686-679X-2023-1-38-49

# Особенности кластера знаний о результативности и востребованности итогов работ российских исследователей в области SIEM-систем

Валерий В. Арутюнов
*Российский государственный гуманитарный университет, Москва, Россия, warut698@yandex.ru*

*Аннотация.* Рассматриваются особенности изменения наукометрических показателей (публикационной активности, цитируемости и востребованности) результатов исследований российских специалистов в 2012–2021 гг. в области SIEM-систем; анализируются выявленные направления исследований в рассматриваемой предметной области, итоги работ по которым отличаются высокой востребованностью. В их числе выделяются: технологии управления информацией и событиями безопасности для защиты компьютерных сетей; анализ методов корреляции событий безопасности в SIEM-системах; система сбора, хранения и обработки информации и событий безопасности на основе средств ELASTIC STACK; подход к разработке SIEM-системы для интернета вещей. Отмечается, что определяемый с учетом индекса Хирша уровень научной активности российских ученых в области SIEM-систем превышает минимальное значение мирового уровня научной активности ученого, равное 16, что свидетельствует о высокой научной квалификации российских исследователей в области SIEM-систем. Определены особенности изменения вышеуказанных показателей, включая экстремумы публикационной активности и цитируемости; выявлены организации России (в основном эти организации из Санкт-Петербурга), чьи итоги исследований в области SIEM-систем активно востребуются другими предприятиями страны. Итоги исследования получены на основе баз данных РИНЦ – Российского индекса научного цитирования.

*Ключевые слова:* SIEM-системы, публикационная активность, управление событиями безопасности, наукометрические показатели, цитируемость, индекс Хирша

*Для цитирования: Арутюнов В.В.* Особенности кластера знаний о результативности и востребованности итогов работ российских исследователей в области SIEM-систем // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 1. С. 38–49. DOI: 10.28995/2686-679X-2023-1-38-49

# Features of the cluster for knowledge about the effectiveness and relevance of the results for the work of Russian researchers in the field of SIEM systems

Valerii V. Arutyunov
*Russian State University for the Humanities, Moscow, Russia,*
*warut698@yandex.ru*

*Abstract*. The article considers features of changes in scientometric indicators (publication activity, citation and demand) of the results for research by Russian specialists in 2012–2021 in the field of SIEM systems; it analyzes the identified research areas in the subject area under consideration, the results of work on which are highly in demand. Among them are: information and security event management technologies for protecting computer networks; analysis of methods for correlating security events in SIEM systems; a system for collecting, storing and processing information and security events based on ELASTIC STACK tools; an approach to developing a SIEM system for the Internet of Things. It is noted that the level of scientific activity of Russian scientists in the field of SIEM systems, determined by taking into account the Hirsch index, exceeds the minimum value of the world level of scientific activity of a scientist equal to 16, which demonstrates a sufficiently high scientific qualification of Russian researchers in the field of SIEM systems. The article defines specifics of changes in the above indicators, including extremes of publication activity and citation as well as the Russian organizations (mainly from St. Petersburg), whose research results in the field of SIEM-systems are actively demanded by other enterprises of the country. The results of the study were obtained on the basis of the RSCI databases – the Russian Science Citation Index.

*Keywords:* SIEM systems, publication activity, security event management, scientometric indicators, citation, Hirsch index

*For citation:* Arutyunov, V.V. (2023), "Features of the cluster for knowledge about the effectiveness and relevance of the results for the work of Russian researchers in the field of SIEM systems", *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 1, pp. 38–49, DOI: 10.28995/2686-679X-2023-1-38-49

*Введение*

С начала XXI в. в России отмечается все возрастающий интерес к вопросам обеспечения информационной безопасности в различных областях экономики и науки, который связан с расширением бизнеса крупных коммерческих фирм и их выходом на международный уровень, развитием страхового и банковского бизнеса, необходимостью ускорения процессов информатизации государственных органов, в том числе вооруженных сил России, повышением уровня криминогенной обстановки в отдельных регионах и рядом других факторов. В связи с этим в XXI в. широкий круг проблем защиты информации в информационно-телекоммуникационных сетях (ИТС) и информационных системах (ИС) находится в центре внимания уже не только специалистов по разработке и эксплуатации этих систем, но и широкого круга пользователей [Арутюнов 2016b].

Развитие информационных и инновационных технологий способствует в наши дни не только повышению уровня комфорта и удобства в работе пользователей и их общению, но и росту количества уязвимостей в работе ИС и ИТС.

В числе основных факторов, способствующих в XXI в. повышению уязвимости обрабатываемой информации, можно выделить следующие:

- резкое расширение круга пользователей, имеющих доступ не только к стационарным компьютерам, но и к различным многофункциональным вычислительным системам;
- развитие телекоммуникационных режимов обработки информации и автоматизация межмашинного обмена данными;
- значительное увеличение объемов информации, собираемой, накапливаемой, хранимой и обрабатываемой с помощью компьютеров и других средств автоматизации;
- рост числа фирм по сравнению с концом XX в., которые обрабатывают информацию, содержащую коммерческую тайну;
- централизация в ИС и ИТС информации многоаспектного назначения, к которой проявляют повышенный интерес не только легитимные пользователи, но и злоумышленники.

Для выявления и устранения уязвимостей и повышения уровня защиты информации используются современные технологии обеспечения ИБ, включающие биометрические системы защиты информации, системы обнаружения и предотвращения вторжений в ИС и ИТС, системы предотвращения утечки информации, SIEM-системы и др.

SIEM-системы включают программные продукты, основное предназначение которых заключается в сборе и анализе информа-

ции о событиях безопасности. Фактически SIEM-системы выполняют задачи управления событиями безопасности (Security Event Management) и управления информацией о безопасности (Security Information Management).

SIEM-системы решают следующие основные задачи:

– отслеживание сигналов тревоги, поступающих от сетевых устройств и приложений;
– обработка полученных данных и определение взаимосвязи между ними;
– выявление отклонений от штатного режима функционирования контролируемых систем;
– оповещение администраторов об обнаруженных инцидентах.

При этом важно понимать, что SIEM-системы не способны предотвращать инциденты, связанные с нарушениями информационной безопасности (ИБ) в организации, и не предназначены для этого.

SIEM-системы используются специалистами службы защиты информации предприятия для выявления на ранних стадиях кибератак и нарушения глобальной или специализированных политик информационной безопасности с целью минимизации возможного ущерба от них. Они позволяют оценить защищенность информационных систем, а также риски, актуальные для организации. В то же время данные, полученные от SIEM-систем, могут быть использованы при расследовании инцидентов и для формирования соответствующей отчетности [Дойникова, Котенко, Чечулин 2015].

SIEM-системы могут обеспечить сбор данных о событиях безопасности четырьмя способами: напрямую с сетевых устройств или с помощью протоколов потоковой передачи данных, с помощью специальных приложений или напрямую из файлов с логами.

В качестве источников информации для SIEM-систем используются:

– системы обнаружения и предотвращения вторжений (IDS/IPS);
– межсетевые экраны;
– системы идентификации и аутентификации;
– антивирусные программные комплексы;
– журналы сетевого оборудования, серверов и рабочих станций;
– системы предотвращения утечки информации (DLP-системы);
– решения для контроля активов и инвентаризации.

О выявленных отклонениях от режима штатного функционирования ИС и ИТС предприятия, который определяется соответствующими политиками безопасности, оповещается администратор безопасности из службы защиты информации предприятия.

Основным потребителем SIEM-систем на сегодня является банковская сфера деятельности, что обусловлено рядом причин. Во-первых, банкам необходимо регулярно проводить аудиты ИБ в соответствии с нормативно-правовыми документами в банковской сфере, а наличие SIEM-систем обусловлено требованиями таких документов, как COBIT, ИСО/МЭК 27001 и др. Во-вторых, банковским организациям проще, чем, например, предприятиям малого или среднего бизнеса обеспечить приобретение, установку и функционирование SIEM-систем, стоимость которых составляет сотни тысяч рублей.

Кроме того, банки работают с финансовой и другой важной конфиденциальной информацией, утечка даже малой части которой может привести к значительным финансовым потерям банковской организации. При этом ежедневно в информационных системах банка происходит значительное количество событий, как правило, важных с точки зрения обеспечения ИБ.

В России за последнее 10 лет отмечается достаточно активное внедрение SIEM-систем, что способствует не только более успешной реализации политик ИБ, но и успешному отслеживанию сбоев в сетевом оборудовании, операционных системах, программном обеспечении (ПО) и т. д. [Графов 2013]. Этому способствовал не только рост объема рынка информационных услуг, но и расширение функциональных возможностей SIEM-систем: только с 2009 по 2014 г. объем российского рынка такого рода систем вырос более чем вдвое – до $ 20 млн.

В Российской Федерации использование SIEM-систем на предприятиях регламентируется рядом нормативно-методических документов (ФЗ РФ № 152 «О персональных данных», нормативно-правовыми документами ФСТЭК России и др.). Этот момент также необходимо учитывать при выборе и последующем использовании SIEM-системы для организации. Важным преимуществом SIEM-систем, кроме анализа информации в режиме реального времени и оповещения об инцидентах ИБ, является тот факт, что доказательства, аккумулированные в базе данных системы, могут выступать в качестве основы не только для разбирательств с нарушителем (злоумышленником), но также могут быть использованы в качестве улик при судебных разбирательствах [Сизов, Киров 2020; Шабуров, Борисов 2016].

В наши дни на российском рынке SIEM-систем в числе наиболее популярных можно отметить следующие: McAfeeESM, IBM QRadar Platform и MaxPatrolSIEM (Positive Technologies).

McAfeeESM (EnterpriseSecurityManager) поставляется в Россию в составе физического и виртуального устройств, а также ПО.

Входящие в состав SIEM-системы три базовых компонента (ESM, Event Receiver и Enterprise Log Manager) могут быть достаточно оперативно развернуты как одно целое или отдельная компонента для распределенных сред.

Платформа QRadar обеспечивает сбор и обработку данных о событиях ИБ из журналов аудита безопасности, а также осуществляет анализ сетевой статистики.

MaxPatrolSIEM является продуктом российской компании Positive Technologies. Он используется в том числе для создания комплекса ГосСОПКА – государственной системы обнаружения, предупреждения и устранения последствий компьютерных атак.

В связи с вышеизложенным возникает вопрос об уровне интенсивности за последние 10 лет научных исследований российских специалистов в области SIEM-систем и их отраженных в публикациях результатах, которые могут представлять интерес для научного сообщества.

## *Наукометрические показатели оценки результатов научных исследований в России в области SIEM-систем*

В России все большее распространение с начала XXI в. получает количественная оценка итогов исследовательской работы научных и образовательных организаций на основе наукометрических показателей результативности работ исследователей (цитируемости *C,* публикационной активности *P* и индекса Хирша *H*).

При этом во многих отраслях наук повышенный интерес для исследователей представляют уже не только указанные наукометрические индексы, но и другие показатели, включая востребованность *V* научным сообществом итогов исследований ученых, определяемую соотношением $C / P$.

Ряд итогов анализа результатов работ российских исследователей в области информационных технологий и по некоторым другим направлениям науки и техники, полученным на основе баз данных РИНЦ [РИНЦ 2022], приводится в широком реестре работ, например: [Арутюнов, Гришина 2020; Arutyunov 2016a; Маршакова-Шайкевич 2008; Grinev 2019] и др.

Ниже анализируется на основе данных РИНЦ динамика изменения в 2012–2021 гг. показателей публикационной активности, определяемой количеством публикаций *P,* их цитируемостью *C* и востребованностью *V* итогов работ российских исследователей в области SIEM-систем.

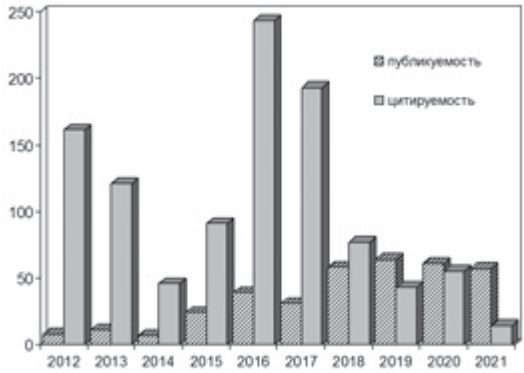


*Рис. 1.* Динамика изменения цитируемости
и публикационной активности российских исследователей
в области SIEM-систем

Динамика изменения в 2012–2021 гг. показателей цитируемости и публикационной активности российских исследователей в области SIEM-систем представлена на рис. 1.

Как видно из рис. 1, наблюдается рост числа публикаций практически до конца анализируемого периода (несмотря даже на кризис 2014 г., когда в динамике изменения публикационной активности был выявлен первый минимум). Этот рост свидетельствует о стабильном интересе российских ученых к анализируемой области исследований практически до конца второго десятилетия XXI в.; при этом в 2018–2021 гг. ежегодное количество публикаций практически стабилизировалось.

Что касается цитируемости, то ее максимум был достигнут в 2016 г., при этом в последующие годы отмечается нелинейное падение значения этого показателя вплоть до минимума в 2021 г.

Динамика изменения востребованности итогов исследований российских ученых за последние 10 лет в области SIEM-систем представлена на рис. 2.

Как следует из рис. 2, максимум показателя отмечался в 2012 г., когда первые появившиеся публикации вызвали в научном сообществе интерес к ним; в дальнейшем наблюдался нелинейный спад $V$ до минимума в 2021 г. в конце анализируемого периода, когда значение $V$ уменьшилось почти в десять раз по сравнению с максимумом. Что касается малых значений индексов цитируемости и востребованности в 2021 г., то этот факт можно объяснить известной причиной: замедленной по целому ряду причин «реакцией» исследователей России на публикации текущего года.

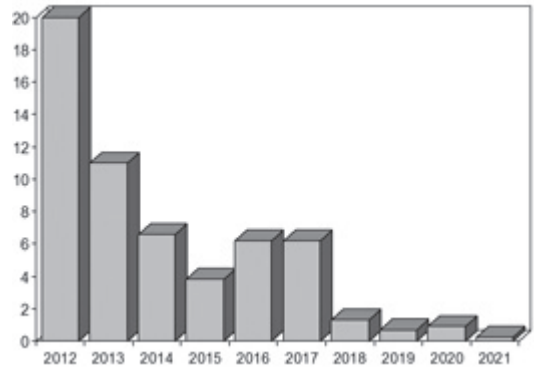


*Рис. 2*. Динамика изменения востребованности итогов работ российских исследователей в области SIEM-систем

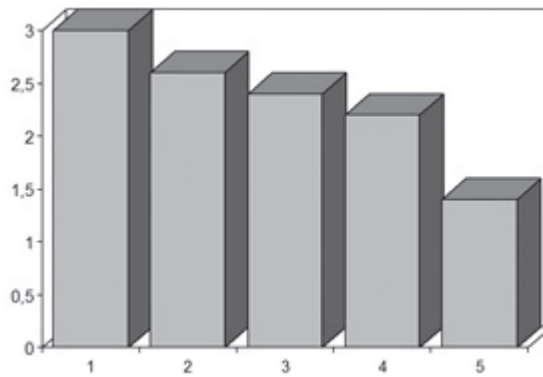


*Рис. 3*. Сравнительный анализ индексов развития результативности исследовательской работы $I_d$ для ряда направлений исследований: 1 – SIEM-системы, 2 – профайлинг, 3 – квантовая криптография, 4 – квантовые технологии, 5 – информационная безопасность

Еще одним показателем, определяющим интенсивность развития научной деятельности в той или иной области наук, является индекс развития результативности исследовательской работы $I_d = I_{2p} / I_{1p}$, где $I_{2p}$ и $I_{1p}$ – суммарная публикационная активность исследователей за последние пять лет ($I_{2p}$) и пять предыдущих лет ($I_{1p}$) [Арутюнов 2017] – в нашем случае за 2017–2021 гг. и 2012–2016 гг. соответственно.

На рис. 3 представлены итоги сравнительного анализа индекса $I_d$ российских ученых для ряда современных направлений исследований: SIEM-системы, профайлинг, квантовая криптография, информационная безопасность, квантовые технологии.

Как видно из рис. 3, наибольшее значение индекса $I_d$ отмечается для SIEM-систем, что свидетельствует о достаточно высокой значимости работ российских ученых в этой области знаний, привлекающей внимание научного сообщества особенно в последние годы по указанным в начале статьи причинам.

Следует отметить, что базы РИНЦ, аккумулируемые в Научной электронной библиотеке России уже более 15 лет, предоставляют исследователям возможность не только определять наукометрические показатели результатов научных исследований. РИНЦ обладает достаточно развитым инструментарием, позволяющим на основе этих показателей определять не только уровень научной активности исследователей в определенной области знаний, но и наиболее востребованные итоги исследований, а также научную результативность исследователей-лидеров (персоналии и организации) в анализируемой области наук.

Так, например, выявленный индекс Хирша для публикаций 2012–2021 гг. в области SIEM-систем равен 17. Это достаточно высокое значение показателя свидетельствует, во-первых, о том, что и в дальнейшем в работе российских ученых можно ожидать стабильную публикационную активность в области SIEM-систем, и, во-вторых, его значение превышает минимум мирового уровня научной активности ученого, равный 16 в соответствии с имеющимися рекомендациями [Ершова 2022]. Этот факт свидетельствует также о достаточно высокой научной квалификации российских исследователей в области SIEM-систем.

В заключение необходимо заметить, что по итогам настоящей работы на основе анализа ежегодного документального потока публикаций в области SIEM-систем в 2012–2021 гг. удалось определить направления исследований в рассматриваемой области знаний, итоги работ по которым отличались высокой цитируемостью и востребованностью. В их число входят: технологии управления информацией и событиями безопасности для защиты компьютерных сетей; подход к разработке SIEM-системы для интернета вещей; система сбора, хранения и обработки информации и событий безопасности на основе средств ELASTIC STACK; анализ методов корреляции событий безопасности в SIEM-системах.

В числе выявленных организаций–лидеров в создании широко востребованных работ в области SIEM-систем вошли три организации из Санкт-Петербурга: СПИИРАН – Институт информатики

и автоматизации РАН, Государственный университет телекоммуникаций им. М.А. Бонч-Бруевича, Политехнический университет Петра Великого.

*Литература*

Арутюнов 2016a – *Арутюнов В.В.* О некоторых результатах приоритетных исследований в области информационной безопасности // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2016. № 2. С. 8–13.

Арутюнов 2016b – *Арутюнов В.В.* Современные проблемы и задачи обеспечения информационной безопасности // Вестник Московского финансово-юридического ун-та МФЮА. 2016. № 2. С. 213–222.

Арутюнов 2017 – *Арутюнов В.В.* Сравнительный анализ результативности научной деятельности в области информационных технологий и защиты информации // Материалы Международной конференции «Информация в современном мире». М.: ВИНИТИ, 2017. С. 13–18.

Арутюнов, Гришина 2020 – *Арутюнов В.В., Гришина Н.В.* Научные кластеры России в области информационных технологий // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2020. № 1. С. 8–24.

Графов 2013 – *Графов А.Ф.* Развитие российских SIEM-систем: SECURITY CAPSULE // Защита информации. Инсайд. 2013. № 5 (53). С. 84–86.

Дойникова, Котенко, Чечулин 2015 – *Дойникова Е.В., Котенко И.В., Чечулин А.А.* Динамическое оценивание защищенности компьютерных сетей в SIEM-системах // Безопасность информационных технологий. 2015. Т. 22. № 3. С. 33–42.

Ершова 2022 – *Ершова С.К.* Инструкция по использованию РИНЦ. URL: https://rfgk.ru/profil-avtora-v-rinc-funkcionalnye-vozmozhnosti-rossiiskii/ (дата обращения 24 февраля 2022).

Маршакова-Шайкевич 2008 – *Маршакова-Шайкевич И.В.* Россия в мировой науке. М.: ИФРАН, 2008. 227 с.

РИНЦ 2022 – *РИНЦ* (Российский индекс научного цитирования). URL: https://elibrary.ru/querybox.asp?scope=newquery (дата обращения 24 февраля 2022).

Сизов, Киров 2020 – *Сизов В.А., Киров А.Д.* Проблемы внедрения SIEM-систем в практику управления информационной безопасностью субъектов экономической деятельности // Открытое образование. 2020. Т. 24. № 1. С. 69–79.

Шабуров, Борисов 2016 – *Шабуров А.С., Борисов В.И.* Разработка модели защиты информации корпоративной сети на основе внедрения SIEM-систем // Вестник Пермского национального исследовательского политехнического ун-та. Электротехника, информационные технологии, системы управления. 2016. № 19. С. 111–124.

Grinev 2019 – *Grinev A.V.* The use of scientometric indicators to evaluate publishing activity in modern Russia // Herald of the Russian Academy of Sciences. 2019. Vol. 89. № 5. P. 451–459.

## *References*

Arutyunov, V.V. (2016a), "About some results of priority research in the field of information security", *Nauchno-tekhnicheskaya informatsiya, Seriya 1. Organizatsiya i metodika informatsionnoi raboty,* no. 2, pp. 8–13.

Arutyunov, V.V. (2016b), "Current issues and tasks of information security", *Vestnik Moskovskogo finansovo-yuridicheskogo universiteta MFUA*, no. 2, pp. 213–222.

Arutyunov, V.V. (2017), "Comparative analysis of the effectiveness of scientific activities in the field of information technology and information security", *Proceedings of the International Conference "Information in the modern world"*, VINITI, Moscow, Russia, pp. 13–18.

Arutyunov, V.V. and Grishina, N.V. (2020), Scientific clusters of Russia in the field of information technology, *RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 1, pp. 8–24.

Doinikova, E.V, Kotenko, I.V. and Chechulin, A.A. (2015), "Dynamic evaluation of the security of computer networks in SIEM systems", *Bezopasnost' informatsionnykh tekhnologii,* vol. 22, no. 3, pp. 33–42.

Ershova, S.K. (2022), "Instructions for using the RSCI", available at: https://rf-gk.ru/profil-avtora-v-rinc-funkcionalnye-vozmozhnosti-rossiiskii/ (Accessed 24 February 2022).

Grafov, A.F. (2013), "Development of Russian SIEM systems: SECURITY CAPSULE", *Information protection. Inside,* no. 5 (53), pp. 84–86.

Grinev, A.V. (2019), "The use of scientometric indicators to evaluate publishing activity in modern Russia", *Herald of the Russian Academy of Sciences*, vol. 89, no. 5, pp. 451–459.

Marshakova-Shaykevich, I.V. (2008), *Rossiya v mirovoi nauke* [Russia in the world science], RAS, Institute of Philosophy, Moscow, Russia.

RSCI (2022), Russian Science Citation Index, available at: https://elibrary.ru/querybox.asp?scope=newquery (Accessed 24 February 2022).

Sizov, V.A. and Kirov, A.L. (2020), "Issues of implementing SIEM systems in practice of the information security management by economic entities", *Otkrytoe obrazovanie,* vol. 24, no. 1, pp. 69–79.

Shaburov, A.S. and Borisov, V.I. (2016), "Development of the information protection model in corporate network based on the implementation of SIEM systems", *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta, seriya Elektrotekhnika, informactsionnye tekhnologii, systemy upravleniya*, no. 19, pp. 111–124.

*Информация об авторе*

*Валерий В. Арутюнов*, доктор технических наук, профессор, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; warut698@yandex.ru

*Information about the author*

*Valerii V. Arutyunov*, Dr. of Sci. (Computer Science), professor, Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Sq., Moscow, Russia, 125047; warut698@yandex.ru

УДК 004.738.5
DOI: 10.28995/2686-679X-2023-1-50-58

# Технологии защиты информации во всемирной сети: патентная аналитика

Ольга В. Маленкова
*Московская международная академия, Москва, Россия, ovm2030@yandex.ru*

Игорь Н. Бычков
*Всероссийская государственная библиотека иностранной литературы им. М.И. Рудомино, Москва, Россия, duke_199090@mail.ru*

*Аннотация.* В статье проведен анализ патентования технологий информационной безопасности в сети Интернет. На сегодняшний день Интернет стал таким же реальным пространством, как наша жизнь. С его помощью мы узнаем новости, погоду, совершаем покупки в магазинах, встречаемся и общаемся с близкими в мессенджерах и соцсетях, посещаем культурно-образовательные мероприятия, получаем дистанционное образование и даже оплачиваемую удаленную работу.

В 2019–2021 гг. всемирная пандемия дала значительный толчок к развитию интернет-технологий и безопасности в сети Интернет. Мир еще больше убедился в ценности интернет-сервисов, и после пандемии многие бизнес-процессы предприятий стали основываться на дистанционных технологиях. Вместе с развитием интернет-сервисов стали активно развиваться безопасные технологии, устраняющие их дестабилизацию.

Статья рассматривает общую ситуацию патентования на момент конца 2021 г. Были определены лидеры и аутсайдеры патентования, анализируются перспективные направления, которые в недалеком будущем станут неотъемлемой частью безопасности в сети Интернет.

В работе рассматривалась база, включающая 938 508 патентных заявок и самих патентов, находящаяся в открытой патентной базе lens.org. Было выявлено общее количество патентных заявок и действующих патентов, приведена динамика патентования, найдены страны-лидеры, определены компании и изобретатели, занимающиеся перспективными разработками в сфере защиты информации в сети Интернет. Выявлены самые перспективные направления в соответствии с международной патентной классификацией.

*Ключевые слова:* Интернет, информационная безопасность, защита информации, патенты, патентная динамика, патентная статистика, программное обеспечение, облачные технологии

*Для цитирования: Маленкова О.В., Бычков И.Н.* Технологии защиты информации во всемирной сети: патентная аналитика // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 1. С. 50–58. DOI: 10.28995/2686-679X-2023-1-50-58

# Technologies for protecting information on the worldwide network. Patent analytics

Ol'ga V. Malenkova
*Moscow International Academy, Moscow, Russia, ovm2030@yandex.ru*

Igor' N. Bychkov
*All-Russia State Library for Foreign Literature, Moscow, Russia, duke_199090@mail.ru*

*Abstract*. The article analyzes the patenting of information security technologies on the Internet. Today, the Internet has become as real a space as our lives. Through it, we learn news, weather, shop in stores, meet and communicate with loved ones in instant messengers and social networks, attend cultural and educational events, get distance education and even paid remote work.

In 2019–2021 the worldwide pandemic has given a big impetus to the development of Internet technologies and Internet security. The world has become even more convinced of the value of Internet services and after the pandemic, many business processes of enterprises began to be based on remote technologies. Together with the development of Internet services, secure technologies that eliminate their destabilization began to actively develop.

The article considers the general patenting situation at the end of 2021. The leaders and outsiders of patenting were identified, promising areas are analyzed, which in the near future will become an integral part of security on the Internet.

The work examined a base including 938,508 patent applications and patents themselves, located in an open patent base lens.org. The total number of patent applications and current patents was revealed, the dynamics of patenting was presented, leading countries were found, companies and inventors engaged in promising developments in the field of information protection on the Internet were identified. The most promising directions were found in accordance with the international patent classification.

*Keywords:* Internet, information security, information protection, patents, patent dynamics, patent statistics, software, cloud technologies

*For citation:* Malenkova, O.V. and Bychkov, I.N. (2023), *"*Technologies for protecting information on the worldwide network. Patent analytics"*, RSUH/RGGU Bulletin. "Information Science. Information Security. Mathematics" Series,* no. 1, pp. 50–58, DOI: 10.28995/2686-679X-2023-1-50-58

## *Введение*

Сеть Интернет – революционное изобретение XX в., вошедшее в жизнь не только отдельных пользователей, но и в сферы науки, культуры, промышленности, в банковскую систему, а также в малый и крупный бизнес.

С распространением Интернета возникли и развивались такие явления, как навязчивый спам, фишинг и т. д., появились компьютерные мошенники, изощренные хакеры, способные манипулировать сознанием и психикой людей, наносящие моральный и экономический ущерб отдельным личностям и доводящие до разорения банки и корпорации.

Изощренные киберпреступления, еще не известные 30 лет назад, к сожалению, не исчезнут и в будущем. Поэтому с развитием компьютерной индустрии проблемы информационной безопасности выходят перед IT-специалистами на первый план.

Решение задач безопасности реализовывалось в двух направлениях:

– в комплексе специальных организационных мероприятий и непосредственно в инженерно-технических изобретениях безопасных программ, технологий и промышленных образцов;
– в юридическом оформлении этих изобретений, т. е. в регистрации патентов.

По международной патентной классификации (МПК) техническая документация систематизируется по двум уровням[1]:

[1] Международная патентная классификация 2021a // ФИПС – Федеральный институт промышленной собственности. URL: https://www1.fips.ru/publication-web/classification/mpk?view=index (дата обращения 1 ноября 2022).

Международная патентная классификация 2021b // Там же. URL: https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=20210101&symbol=none&menulang=en&lang=en&viewmode=f&fipcpc=

– базовый, с основными классами;
– расширенный, с декомпозицией на классы и подклассы.

### *Наиболее патентуемые технологии по классификации МПК*

В табл. 1 для примера представлена небольшая часть базовых тематических классов–лидеров в патентовании (примерно ⅓ от всего количества).

*Таблица 1*

Значение МПК классов

| № | Класс МПК | Значение |
|---|---|---|
| 1 | H04L29/06 | Методы безопасной передачи информации по каналам связи в сети |
| 2 | H04L29/08 | Процедуры управления передачей информации в сети, например уровнем данных в канале передачи |
| 3 | G06F17/30 | ИБ в области управления контентом сайта |
| 4 | H04L9/32 | Устройство секретной связи со средствами для установления личности или полномочий пользователя системы |
| 5 | G06F15/16 | ИБ в сфере балансировки нагрузки для ПО |

### *Динамика патентования в сфере информационной безопасности*

Рассмотрим динамику патентования по годам (рис. 1).

Из графика видно, что спрос на безопасные технологии плавно возрастал с 1997 г., когда Интернет набирал силу и популярность.

Первые 11 патентов были зарегистрированы в 1989 г. Застой в патентовании в 2008–2009 гг. и в 2014–2016 гг. объясняется мировым финансовым кризисом.

---

no&showdeleted=yes&indexes=no&headings=yes&notes=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart (дата обращения 1 ноября 2022). Патентная база lens.org 2021. URL: https://www.lens.org/ (дата обращения 1 ноября 2022).

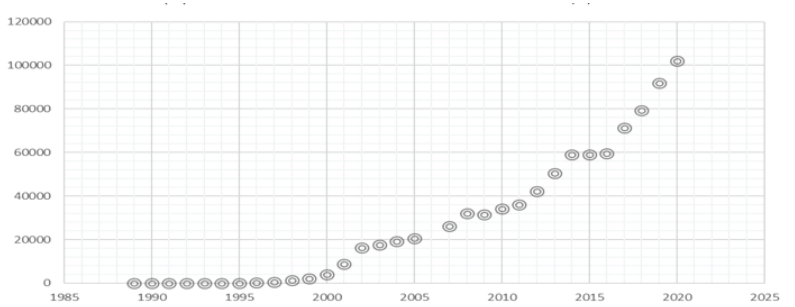


*Рис. 1.* Динамика патентования в области ИБ

## *География, изобретатели и правообладатели патентной собственности*

Подача большего количества заявок и регистрация патентов происходит в США (US); также заявки и патенты регистрируются через процедуру Международной организации интеллектуальной собственности (WIPO – WO) или Европейской патентной организации (EPO – EP), а также патентными организациями многих стран мира (рис. 2).

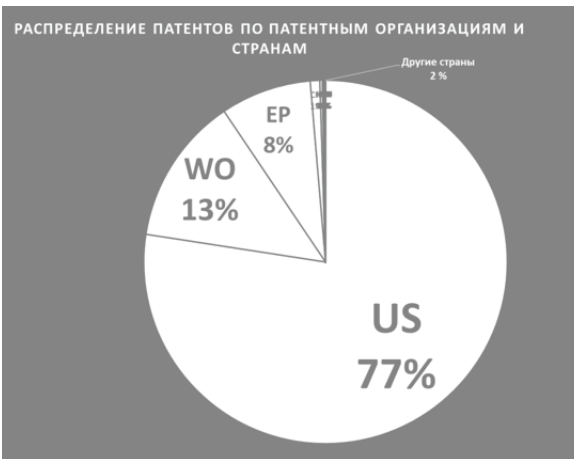


*Рис. 2.* Распределение патентования в США и других патентных организациях

Чаще всего заявителями являются компании с представительствами в различных странах, а также отдельные изобретатели.

Среди крупных компаний, возглавляющих список патентообладателей, можно отметить изобретения разработок компьютерной архитектуры (IBM, Intel, Qualcomm, Apple), виртуализации (Citrix, Systems, Huawei tech), телефонии (Ericsson telefon, Huawei tech, Nokia, Sony, Blackberry), сетевых технологий (Cisco, Hewlet-Packard, Huawei, At & IP), облачных сервисов (Google, Salesforce); также к лидерам патентования относятся поставщики интернет-сервисов (Amazon Tech, Google, Facebook) и компании–владельцы платежных систем (Visa и Mastercard).

На диаграмме (рис. 3) представлены изобретатели с наибольшим количеством патентов. Среди них лидируют выходцы из США (один иммигрировал из Канады и несколько из Китая).

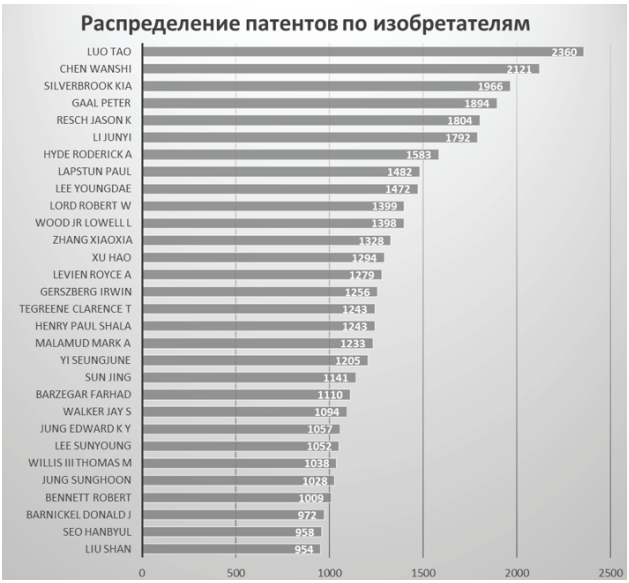


*Рис. 3.* Изобретатели-лидеры

Двумя процентами заявок владеют отдельные изобретатели и правообладатели, оформившие свои разработки в патентных организациях, кроме США, WIPO и EPO.

Лидерами в изобретательстве безопасных технологий являются: Китай, Народная республика Корея, Япония, Франция, Канада, Англия, Австралия, Германия и Россия (рис. 4).

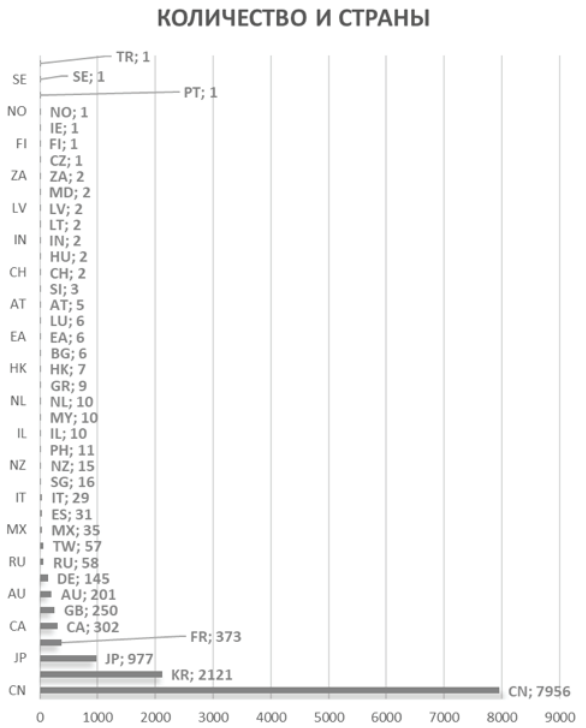


*Рис. 4.* Распределение патентов по странам

## *Заключение*

В настоящее время (конец 2021 г.) суммарное фондовое значение заявок (63%) и патентов (37%) составляет величину 938 508[2].

Из рис. 5 видно, что такое процентное соотношение между заявками и оформленными патентами свидетельствует о взрывном интересе к безопасным технологиям.

Совершенно очевидно, что такому положению вещей способствовала в определенной мере коронавирусная пандемия, продемонстрировав, с одной стороны, огромные возможности Интернета, а с другой стороны, незащищенность пользователей от мошенников и хакеров [Гришина, Маленкова, Бычков 2017].

[2] Патентная база lens.org 2021. URL: https://www.lens.org/ (дата обращения 1 ноября 2022).

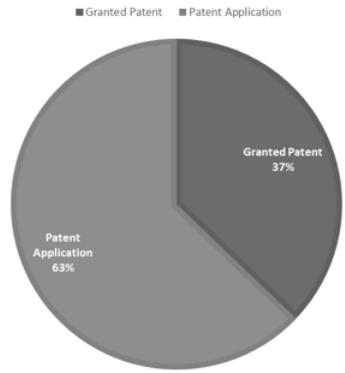


*Рис. 5.* Состояние патентования
в сфере безопасности Интернета

Статистика регистрирует рост киберпреступлений в платежных банковских и торговых системах, в том числе при краже и сбыте конфиденциальной информации и т. д. В связи с этими событиями в последнее время наиболее перспективные направления указаны в табл. 2.

*Таблица 2*

Значение МПК классов

| № | Класс МПК | Значение |
|---|---|---|
| 1 | G06Q30/02 | Системы, защищающие профиль пользователя от маркетинговых исследований |
| 2 | G06K9/00 | Биометрия |
| 3 | H04L9/08 | Устройства для конфиденциальной связи с ключевым распределением |
| 4 | H04W12/06 | Устройства аутентификации в беспроводных сетях |
| 5 | H04W72/04 | Методы, основанные на запланированном доступе в беспроводную сеть |
| 6 | G06N20/00 | Компьютерные системы машинного обучения в сфере ИБ |
| 7 | H04W4/80 | Защита данных для радиоустройств короткого радиуса действия (NFC, RFID и т. п.) |
| 8 | H04W12/00 | Устройства для обеспечения конфиденциальной связи в беспроводных сетях |

## *Литература*

Гришина, Маленкова, Бычков 2017 – *Гришина Н.В., Маленкова О.В., Бычков И.Н.* Проблемы обеспечения информационной безопасности при использовании облачных технологий // Современные проблемы и задачи обеспечения информационной безопасности СИБ – 2017: Международная научно-практическая конференция (г. Москва, 18 апреля 2017 г.): Сб. статей. М.: Московский финансово-юридический университет МФЮА, 2017. С. 187–191.

## *References*

Grishina, N.V., Malenkova, O.V., and Bychkov I.N. (2017), "Issues of ensuring information security at the use of cloud computing", *Modern issues and challenges of information security ISS – 2017. International Scientific and Practical Conference (Moscow, April 18, 2017)". Coll. of articles*, Financial and Law University MFUA, Moscow, Russia, pp. 187–191.

## *Информация об авторах*

*Ольга В. Маленкова*, кандидат технических наук, доцент, Московская международная академия, Москва, Россия; 129075, Россия, Москва, ул. Новомосковская, д. 15 А, стр. 1; ovm2030@yandex.ru

*Игорь Н. Бычков,* Всероссийская государственная библиотека иностранной литературы им. М.И. Рудомино, Москва, Россия; 109240, Россия, Москва, ул. Николоямская, д. 1; duke_199090@mail.ru

## *Information about the authors*

*Ol'ga V. Malenkova*, Cand. Of Sci. (Mechanical Engineering), associate professor, Moscow International Academy, Moscow, Russia; bld. 15 A/1, Novomoskovskaya Str., Moscow, Russia, 125075; ovm2030@yandex.ru

*Igor' N. Bychkov*, All-Russia State Library for Foreign Literature, Moscow, Russia; bld. 1, Nikoloyamskaya Str., Moscow, Russia, 109240; duke_199090@mail.ru

УДК 519.6+159.943
DOI: 10.28995/2686-679X-2023-1-59-72

# Проблема автоматизированной детекции виктимного поведения индивида с точки зрения качества сервиса

Анна Б. Клименко
*Российский государственный гуманитарный университет, Москва, Россия, anna_klimenko@mail.ru*

*Аннотация*. В данной статье рассмотрен вопрос автоматической детекции потенциально виктимного поведения индивида в цифровой среде в аспекте качества предоставляемого сервиса. Анализ прецедентов обнаружения виктимного поведения демонстрирует связь между наличием у индивида множественных ложных аккаунтов и высоким потенциалом виктимного поведения. Такой вывод позволяет обнаруживать индивидов, склонных к виктимизации, через обнаружение их ложных аккаунтов; также анализ современных работ в данной предметной области демонстрирует, что одним из наиболее эффективных методов обнаружения ложных псевдонимов в сети является стилометрический анализ, который, однако, является самым трудоемким и требующим анализа наибольшего количества данных. В связи с этим актуализируется проблема соответствия качества сервисов заявленному уровню с точки зрения времени выполнения операций, что является нетривиальной задачей при поиске и обработке информации в сильно распределенных цифровых средах. В данной статье представлены подходы и методы сокращения времени выполнения операций поиска и детекции индивидов с виктимным поведением за счет адаптации параметров системы к условиям поиска, распараллеливания и размещения обработки данных поблизости от их источников, а также приведены экспериментальные расчеты, на основе которых сделаны выводы о целесообразности использования адаптивных свойств систем и элементов концепции туманных вычислений.

*Ключевые слова:* распределенные вычисления, стилометрический анализ, виктимное поведение, цифровое пространство

*Для цитирования: Клименко А.Б.* Проблема автоматизированной детекции виктимного поведения индивида с точки зрения качества сервиса // Вестник РГГУ. Серия «Информатика. Информационная безопасность. Математика». 2023. № 1. С. 59–72. DOI: 10.28995/2686-679X-2023-1-59-72

# An issue of automatic individual's victim behavior detection from the quality of service point of view

Anna B. Klimenko
*Russian State University for the Humanities, Moscow, Russia, anna_klimenko@mail.ru*

*Abstract*. This article discusses the issue of automatic detection of potentially victimized behavior of an individual in a digital environment in terms of the quality of the service provided. An analysis of precedents for detecting the victim behavior demonstrates a connection between the presence of multiple false accounts in an individual and a high potential for the victim behavior. Such an inference makes it possible to detect individuals prone to victimization through the detection of their false accounts. Also, the analysis of recent works in the subject area demonstrates that one of the most effective methods for detecting false aliases in the network is stylometric analysis, which, however, is the most time-consuming and requires the analysis of the largest amount of data. In this regard, the issue of compliance with the quality of services to the declared level in terms of the execution time of operations becomes relevant, which is a non-trivial task when searching and processing information in highly distributed digital environments. The article presents approaches and methods for reducing the execution time of search operations and detection of individuals with the victim behavior by adapting system parameters to search conditions, as well as by parallelizing and placing data processing close to their sources. Experimental calculations are also presented, on the basis of which conclusions are drawn about the appropriateness of using the adaptive properties of systems and elements of the fog computing concept.

*Keywords:* distributed computing, stylometric analysis, victimized behavior, digital space

*For citation:* Klimenko, A.B. (2023), "An issue of automatic individual's victim behavior detection from the quality of service point of view", *RSUH/RGGU Bulletin. "Informatics. Information security. Mathematics" Series*, no. 1, pp. 59–72, DOI: 10.28995/2686-679X-2023-1-59-72

## *Введение*

В настоящее время проблема виктимного поведения индивидов приобрела особую актуальность в условиях расширения и совершенствования ряда коммуникационных средств, в том числе наличия среды информационного пространства Интернет как платформы, обеспечивающей практически ничем не ограниченное общение.

Информационное пространство сети Интернет обладает такими особенностями, как физическая непредставленность индивида и в большинстве случаев отсутствие ответственности за совершенные поступки, что, в свою очередь, продуцирует практически идеальные условия для различных прецедентов взаимодействия по типу «жертва-злоумышленник».

Виктимное поведение определяется как поведение, результатом которого становится преступление [Melnik, Korovin, Klimenko 2019; Невский, Кулакова 2007].

До появления и распространения сети Интернет виктимология предоставляла проработанные методы детекции и прогнозирования виктимного поведения индивида, однако в цифровую эру сложилась ситуация, когда данная дисциплина уже не располагает методологией, ориентированной на использование в цифровом пространстве.

Таким образом, актуальной становится задача создания автоматических или автоматизированных средств, которые бы позволяли по определенному набору признаков производить детекцию индивидов, склонных – либо реализующих – виктимное поведение в сети.

Следует отметить, что при создании такой информационной системы возникает ряд проблем, связанных с таким понятием, как качество предоставляемого сервиса (Quality of service, QoS). Признаки виктимного поведения, как правило, сокрыты в текстах либо изображениях и видео, которые являются данными, распределенными по цифровому пространству сети Интернет. Соответственно, их сбор и обработка могут стать неприемлемо долгими, в то время как индивид, пользующийся сервисом поиска/детекции виктимного поведения, либо индивид, желающий каким-то образом фильтровать контент своих новостных лент от угроз, не пожелает ждать результатов слишком долго.

По этой причине основной проблемой, рассматриваемой в данной статье, является проблема сокращения времени, необходимого для систем поиска и детекции проявлений виктимного поведения в сети.

## *1. Виктимное поведение и ложные псевдонимы*

Исследования виктимного поведения индивида в информационном пространстве показали, что такое поведение в большинстве случаев сопровождается либо наличием у индивида фейковых аккаунтов, либо наличием некоторого числа фейковых аккаунтов в информационном окружении индивида.

Фейковый аккаунт (fake account) – это аккаунт, в котором индивид делает вид, что является кем-то другим.

Такие аккаунты используются жертвами кибербуллинга с целью восстановить свой рейтинг в рамках определенных сообществ, либо, наоборот, используются индивидами с целью реализации активного виктимного поведения, включая публикацию материала, содержащего личные сведения.

Попытки борьбы с виктимным поведением пользователей в настоящее время могут быть разделены на три основные стратегии, в рамках которых разрабатываются технические средства [Chen, Conroy, Rubin 2015, pp. 15–19; Markines, Cattuto, Menczer 2009, pp. 41–48]:

- средства родительского контроля (фильтры, централизованные БД «черных списков» и т. п.);
- административные методы (крупные социальные сети пытаются бороться с фейками, деструктивным контентом);
- сторонние программные средства.

Фейковые аккаунты используются и жертвами как в попытке избавиться от преследования, так и в попытке получить доступ к закрытой информации, принимать участие в разного рода сообществах деструктивной направленности, а также и самим принимать участие в травле других пользователей сети.

Таким образом, опираясь на проведенное исследование и на наличие прецедентов использования множественных фейк-аккаунтов для реализации частных случаев виктимного поведения, методом обобщения получим новый признак, способствующий идентификации виктимного поведения в сети – наличие у индивида множественных аккаунтов-двойников.

## *2. Обзор известных методов обнаружения псевдонимов-«двойников»*

Обзор публикаций в данной области показал, что к настоящему времени сложились следующие основные направления обнаружения «двойников»:

– методы на основании схожести псевдонимов (String-based matching);
– методы на основе стилометрии (Stylometric matching);
– методы на основе анализа временного профиля (Time profile-based matching);
– методы на основе анализа социального взаимодействия (Social network-based matching).

*Методы на основе схожести псевдонимов*. Пользовательские псевдонимы, как правило, являются некой текстовой строкой. Зачастую пользователи создают похожие псевдонимы в различных сетях, по крайней мере, этот случай работает тогда, когда пользователь не пытается скрываться или вести противозаконную деятельность [Shaikh, Memon, Wiil 2011, pp. 216–219]. Предложено несколько вариантов проверки схожести псевдонимов, например: [Levenshtein 1966].

Однако эта стратегия не будет работать, когда пользователь умышленно скрывается под другим псевдонимом.

*Методы на основе стилометрии*. Под стилометрией понимают статистический анализ текста [Zheng, Li, Chen, Huang 2006, pp. 378–393]. К настоящему времени существует достаточно широкий круг работ, посвященный идентификации авторства путем анализа текста, например: [Stamatatos 2009, pp. 538–556; Abbasi, Chen 2008, pp. 7:1–7:29; Juola 2006, pp. 233–334].

Следует отметить, что большинство работ в данной области посвящены проблеме идентификации автора малой размерности (когда количество кандидатов в авторы относительно невелико), и гораздо меньшее количество работ посвящено идентификации авторов в больших киберпространствах либо, наоборот, с использованием малого количества материала для создания вектора характеристик текстов. Методы на основе стилометрии считаются достаточно надежными и перспективными [Ho, Ng 2016].

*Методы на основе анализа временного профиля*. Методы данной группы опираются на допущение о том, что не может один и тот же индивид публиковать пост в одно и то же время. На основе этой гипотезы составляются временные профили пользователей, причем могут включать не только публикацию постов, но и прочие действия. Примерами формирования временных профилей являются работы [Johansson, Kaati, Shrestha 2013].

*Методы на основе социального взаимодействия*. Эта группа методов опирается на гипотезу о том, что псевдонимы одного и того же индивида в сети с высокой вероятностью взаимодействуют с одними и теми же аккаунтами. Тема затрагивается в работе [Shres-

tha, Atig, Kaati, Cassel 2014]. Однако проблемой является и то, что не всегда доступны списки друзей пользователя в Интернете, и по этой причине возможны варианты построения сети связей: на основе сообщений, на основе бесед и т. п.

Следует отметить, что методы на основе стилометрии считаются наиболее точными, однако в то же время являются и наиболее трудозатратными. Это актуализирует вопрос о методах сокращения времени работы таких систем.

### *3. Способы снижения ресурсных затрат при детекции ложных псевдонимов*

В самом общем виде составляющие компоненты методологии построения сервиса базируются на основных требованиях к его функциональной нагрузке, то есть:

- в функционал должны входить средства формирования векторов, описывающих авторский стиль;
- должны входить средства поиска и сравнения векторов в интернет-пространстве;
- в него также должны входить средства, позволяющие адаптироваться к изменениям внешней среды.

Здесь адаптивное поведение определяет возможность корректировать пространство поиска фейковых псевдонимов в цифровом пространстве на основе интеллектуального анализа истории поисковых операций, а также способность корректировать величину текстовой выборки, необходимой для построения вектора текстовых характеристик. Это позволяет, во-первых, ускорить обнаружение фейк-аккаунтов по причине частого пересечения множества пользователей определенных ресурсов [Sadia, Aylin, Stolerman, Greenstadt, McCoy 2014, pp. 212–226] и улучшить качество результатов поиска, исключая необходимость повторного поиска, и, во-вторых, уменьшить область поиска в случае, когда это возможно.

Использование концепции туманных вычислений позволяет достигнуть следующих позитивных эффектов:

- за счет разгрузки сетевой инфраструктуры снизить время реакции систем на внешние воздействия;
- достижима возможность разгрузки краевых, пользовательских устройств за счет возможности перенести вычисления в «туманный слой» [Moysiadis, Sarigiannidis, Moscholios 2018];
- достижима разгрузка серверного оборудования за счет смещения вычислительной нагрузки в «туманный» слой;

– возможно предоставление ресурсов для сложных пользовательских задач, а также для хранения данных за счет «туманного» слоя.

Размещение данных в «туманном» слое также оказывает положительный эффект на быстродействие системы и ее масштабируемость, поскольку сервер централизованной архитектуры является «узким местом» системы, ухудшает масштабируемость и в случае загруженности негативно влияет на временные характеристики систем.

Еще одним современным трендом в построении распределенных систем является переход к децентрализованным системам с равноправными узлами, а также к распределенному хранению данных по типу распределенного реестра.

Внедрение технологий распределенного реестра также целесообразно в случае необходимости улучшения масштабируемости и времени реакции сервиса.

На рис. 1 показаны наиболее трудозатратные операции по поиску и детекции ложных аккаунтов, а также схематично показаны базовые подходы для сокращения ресурсопотребления.

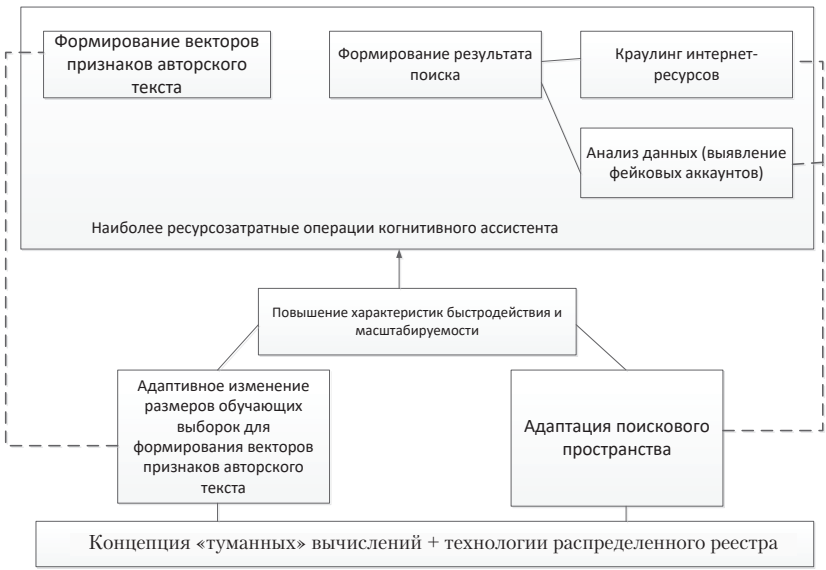


*Рис. 1.* Подход к повышению быстродействия и масштабируемости системы

## *4. Адаптация размеров обучающих выборок*

В качестве примера применения предлагаемого общего подхода к сокращению времени выполнения системой операций оценим влияние вариативных размеров выборок авторских текстов на скорость работы системы.

Размеры обрабатываемых текстовых выборок могут варьироваться в зависимости от типов источников данных.

Пусть $c$ – количество характеристик (размерность вектора), $i$ – номер характеристики. Тогда трудоемкость обработки текста для получения $i$-й характеристики будет равна:

$$x_i = \xi_i V_t,$$

где $V_t$ – объем текстовой выборки (слов), $\xi_i$ – коэффициент пропорциональности, определяющий трудоемкость обработки текста для получения $x_i$.

В этом случае трудоемкость получения вектора текстовых характеристик будет определяться следующим образом:

$$X = \sum_{i=1}^{c} x_i = \sum_{i=1}^{c} \mathsf{x}_i V_t, \qquad (1)$$

Время получения вектора, соответственно, определяется как:

$$T_c = \alpha X,$$

где $\alpha$ – коэффициент пропорциональности, определяемый производительностью узла, на котором производится обработка.

Полное время получения векторных характеристик суммируется из следующих слагаемых, считая текстовую выборку уже собранной:

– время доставки выборки на устройство, где будет осуществляться формирование вектора текстовых характеристик;
– собственно время получения вектора;
– время доставки результата пользователю.

Данную процедуру имеет смысл представлять, принимая во внимание расположение узла, осуществляющего вычисление вектора, относительно источника текстовой выборки, как это отражено в формуле 2:

$$T_0 = V_t k + T_c + k\eta c, \quad (2)$$

где $T_0$ – полное время получения пользователем вектора текстовых признаков по сформированной текстовой выборке;

$k$ – коэффициент, определяющий скорость передачи по сети в зависимости от количества транзиторных участков;

$\eta c$ – объем вектора текстовых характеристик.

Зависимость времени получения пользователем вектора текстовых признаков от объема текстовой выборки может быть оценена и имеет вид графика на рис. 2.


*Рис. 2.* Зависимость времени получения вектора текстовых характеристик (в ед. моделирования) от объема текстовой выборки

Результаты оценивания зависимости времени выполнения пользовательской процедуры от среднего объема текстовой выборки позволяют рассматривать предлагаемый подход как перспективный в аспекте поддержания должного уровня QoS.

## *5. Адаптация областей поиска*

Рассмотрим далее эффект влияния адаптивности системы по области поиска. Определим следующие переменные:

$T_s$ – время одного сеанса поиска, где для имеющегося вектора текстовых характеристик производится поиск возможных псевдонимов-двойников;

$W_i$ – объем информации для краулинга, т. е. фактически объем ресурсов, который необходимо считать с ресурса;
$\mu$ – скорость работы краулера;
σ – скорость поиска псевдонимов по полученным от краулера данным, включая составление векторов текстовых характеристик для потенциальных псевдонимов.

Тогда одна сессия поиска ложных псевдонимов для заданного вектора текстовых характеристик на $i$-м ресурсе составит:

$$T_{si} = \mu W_i + \sigma P_i + \theta(P_i), \quad (3)$$

где $P_i$ – общее количество псевдонимов, содержащихся на $i$-м ресурсе,
$\theta(P_i)$ – время сравнения векторов, подразумевая сравнение эталонного вектора со всеми сформированными на основе обнаруженных псевдонимов.

Таким образом, для $m$ обрабатываемых ресурсов получаем следующее выражение:

$$T_s = \sum_{i=1}^{m} \mu W_i + \sigma P_i + \theta(P_i)\phi. \quad (4)$$

Полученная оценка имеет вид, как показано на графике рис. 3.

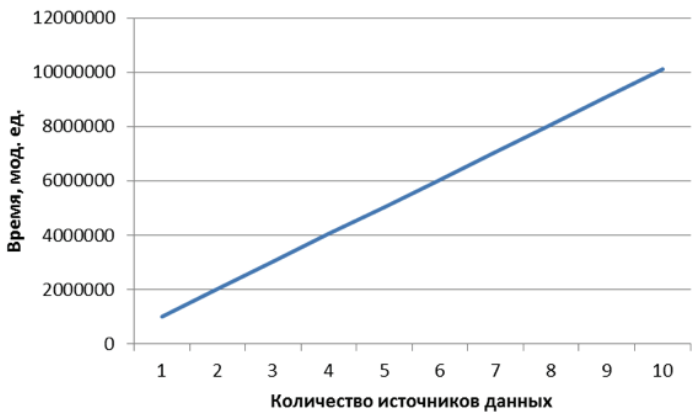


*Рис. 3.* Зависимость времени поиска от количества исследуемых ресурсов

То есть видно, что возможность адаптации поискового пространства значительно влияет на время получения ответа от системы и возможность сужения области выбора ресурсов для поиска может существенно сократить время работы системы.

## *Заключение*

В статье рассмотрена проблема временных затрат при функционировании систем детекции виктимного поведения в сети на примере системы, реализующей стилометрический анализ. Наиболее ресурсозатратными процедурами с точки зрения времени являются: формирование векторов признаков авторского текста, сбор данных и анализ данных (сравнение всех сформированных векторов признаков текста с референсным).

Предлагаемый подход к уменьшению временных затрат включает: параллельный сбор данных, использование технологий распределенного реестра, а также адаптацию пространства поиска ложных псевдонимов, в случае если фиксируется пересечение аудиторий ресурсов. Таким образом достигается улучшение показателей QoS системы автоматической детекции виктимного поведения индивида в сети.

## *Литература*

Невский, Кулакова 2007 – *Невский Н.Н., Кулакова А.А.* Виктимология и ее основополагающие аспекты: Монография. Владимир: Владимирский юридический ин-т Федеральной службы исполнения наказаний, 2007.

Abbasi, Chen 2008 – *Abbasi A., Chen H.* Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace // ACM Trans. Inf. Syst. 2008. Vol. 26, no. 2. P. 7:1–7:29.

Chen, Conroy, Rubin 2015 – *Chen Y., Conroy N.J., Rubin V.L.* Misleading online content: Recognizing clickbait as false news // Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection – WMDD'15, November 13, 2015, Seattle, WA, USA. New York: ACM, 2015. P. 15–19.

Ho, Ng 2016 – *Ho T.N., Ng W.K.* Application of Stylometry to DarkWeb Forum User Identification // Information and Communications Security – ICICS 2016 / K.Y. Lam, C.H. Chi, S. Qing (eds.). Cham: Springer, 2016. P. 173–183 (Lecture Notes in Computer Science. Vol. 977).

Johansson, Kaati, Shrestha 2013 – *Johansson F., Kaati L., Shrestha A.* Detecting Multiple Aliases in Social Media // The International Symposium on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI 2013),

Niagara Falls, ON, Canada, 26–27 August 2013. New York, NY: IEEE, 2013. P. 1004–1011.

Juola 2006 – *Juola P.* Authorship attribution // Found. Trends Inf. Retr. 2006. Vol. 1, no. 3. P. 233–334.

Levenshtein 1966 – *Levenshtein V.* Binary Codes Capable of Correcting Deletions, Inser-tions and Reversals // Soviet Physics Doklady, 1966. Vol. 10. P. 707–710.

Markines, Cattuto, Menczer 2009 – *Markines B., Cattuto C., Menczer F.* Social spam detection // Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web. New York: ACM, 2009. P. 41–48.

Melnik, Korovin, Klimenko 2019 – *Melnik E., Korovin I., Klimenko A.* A Cognitive Assistant Functional Model and Architecture for the Social Media Victim Behavior Prevention // Artificial Intelligence Methods in Intelligent Algorithms (CSOC-2019). Cham: Springer, 2019. P. 51–61 (Advances in Intelligent Systems and Computing. Vol. 985).

Moysiadis, Sarigiannidis, Moscholios 2018 – *Moysiadis V., Sarigiannidis P., Moscholios I.* Towards Distributed Data Management in Fog Computing // Wireless Communications and Mobile Computing. 2018. № 1. P. 1–14.

Sadia, Aylin, Stolerman, Greenstadt, McCoy 2014 – *Sadia A., Aylin C.I., Stolerman A., Greenstadt R., McCoy D.* Doppelgänger Finder: Taking Stylometry to the Underground // Proceedings of the 2014 IEEE Symposium on Security and Privacy, May 18–21, 2014, San Jose, California, USA. New York, NY: IEEE, 2014. P. 212–226.

Shaikh, Memon, Wiil 2011 – *Shaikh M., Memon N., Wiil U.* Extended approximate string matching algorithms to detect name aliases // Proceedings of the 2011 IEEE International Conference on Intelligence and Security Informatics, July 2011. New York, NY: IEEE, 2011. P. 216 –219.

Shrestha, Atig, Kaati, Cassel 2014 – *Shrestha A., Atig, M.F., Kaati L., Cassel S.* Activity Profiles in Online Social Media // 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 17–20 August 2014, Beijing, China. New York, NY: IEEE, 2014.

Stamatatos 2009 – *Stamatatos E.* A survey of modern authorship attribution methods // Journal of the American Society for Information Science and Technology. 2009. Vol. 60, no. 3. P. 538–556.

Zheng, Li, Chen, Huang 2006 – *Zheng R., Li J., Chen H., Huang Z.* A framework for authorship identification of online messages: Writing-style features and classification techniques // J. Am. Soc. Inf. Sci. Technol. 2006. Vol. 57, no. 3. P. 378–393.

## *References*

Abbasi, A. and Chen, H. (2008), “Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace”, *ACM Trans. Inf. Syst.*, vol. 26, no. 2, pp. 7:1–7:29.

Chen, Y., Conroy, N.J., and Rubin, V.L. (2015), "Misleading online content: Recognizing clickbait as false news", *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection – WMDD'15*, November 13, 2015, Seattle, WA, USA, ACM, New York, USA, pp. 15–19.

Ho, T.N. and Ng, W.K. (2016), "Application of Stylometry to DarkWeb Forum User Identification", in: Lam, K.Y., Chi, C.H., Qing, S. (eds.), *Information and Communications Security – ICICS 2016*, Springer, Cham, Switzerland, pp. 173–183. (*Lecture Notes in Computer Science*, vol. 9977)

Johansson, F., Kaati, L. and Shrestha, A. (2013), "Detecting Multiple Aliases in Social Media", *Proceedings of the 2013 International Symposium on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI 2013)*, Niagara Falls, ON, Canada, 26–27 August 2013), IEEE, New York, NY, USA, pp. 1004–1011.

Juola, P., (2006), "Authorship attribution", *Found. Trends Inf. Retr.*, vol. 1, no. 3, pp. 233–334.

Levenshtein, V. (1966), "Binary Codes Capable of Correcting Deletions", *Insertions and Reversals. Soviet Physics Doklady*, vol. 10, pp. 707–710.

Markines, B., Cattuto, C., and Menczer, F. (2009), "Social spam detection", *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web,* ACM, New York, USA, pp. 41–48.

Melnik, E., Korovin, I. and Klimenko, A. (2019), "A Cognitive Assistant Functional Model and Architecture for the Social Media Victim Behavior Prevention", *Artificial Intelligence Methods in Intelligent Algorithms. CSOC 2019,* Springer, Cham, Switzerland, pp. 51–61 (Advances in Intelligent Systems and Computing, vol. 985).

Moysiadis, V., Sarigiannidis, P., and Moscholios, I. (2018), "Towards Distributed Data Management in Fog Computing", *Wireless Communications and Mobile Computing*, vol. 1. pp. 1–14.

Nevskiy, N. and Kulakova, A. (2007), *Viktimologiya i ee osnovopolagayushhie aspekty: monografiya* [Victimology and its basic aspects. Monograph], VLI of FPS of Russia, Vladimir, Russia, 180 p.

Sadia, A., Aylin, C.I., Stolerman, A., Greenstadt, R. and McCoy, D. (2014), "Doppelgänger Finder: Taking Stylometry to the Underground", *Proceedings of the 2014 IEEE Symposium on Security and Privacy*, May 18–21, 2014, San Jose, California, USA, IEEE, New York, NY, USA, pp. 212–226.

Shaikh, M., Memon, N., and Wiil, U. (2011), "Extended approximate string matching algorithms to detect name aliases", *Proceedings of the 2011IEEE International Conference on Intelligence and Security Informatics,* July 2011, IEEE, New York, NY, USA, pp. 216–219.

Shrestha, A., Atig, M.F., Kaati, L. and Cassel, S. (2014), "Activity Profiles in Online Social Media", *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, 17–20 August 2014, Beijing, China, IEEE, New York, NY, USA, 2014.

Stamatatos, E. (2009), "A survey of modern authorship attribution methods", *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, pp. 538–556.

Zheng, R., Li, J., Chen, H., and Huang, Z. (2006), "A framework for authorship identification of online messages: Writing-style features and classification techniques", *J. Am. Soc. Inf. Sci. Technol.*, vol. 57, no. 3, pp. 378–393.

### *Информация об авторе*

*Анна Б. Клименко*, кандидат технических наук, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; anna_klimenko@mail.ru.

### *Information about the author*

*Anna B. Klimenko*, Cand. of Sci. (Computer Science), Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Sq., Moscow, Russia, 125047; anna_klimenko@mail.ru.

Дизайн обложки
*Е.В. Амосова*

Корректор
*Н.В. Москвина*

Компьютерная верстка
*Н.В. Москвина*

Подписано в печать 10.03.2023.
Формат 60×90 1/16.
Уч.-изд. л. 4,5. Усл. печ. л. 4,6.
Тираж 1050 экз. Заказ № 1689

Издательский центр
Российского государственного
гуманитарного университета
125047, Москва, Миусская пл., 6
www.rsuh.ru