

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года
Журнал выходит ежеквартально

Учредитель:
**Федеральный исследовательский центр
«Информатика и управление» Российской академии наук**

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета
академик РАН Г. И. Савин
академик РАН А. Л. Стемпковский
член-корреспондент РАН Ю. Б. Зубарев
профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)
профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)
профессор В. Ротарь (V. Rotar, San-Diego, USA)
профессор М. Финкельштейн (M. Finkelstein, Bloemfontein, South Africa)
профессор В. Хофкирхнер (W. Hofkirchner, Wien, Austria)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор
профессор, д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора
д.т.н. В. Н. Захаров
проф., д.ф.-м.н. А. И. Зейфман
проф., д.т.н. В. Д. Ильин
проф., д.т.н. К. К. Колин
проф., д.ф.-м.н. В. Ю. Королев
д.ф.-м.н. Р. В. Разумчик
д.ф.-м.н. В. И. Сеницын
проф., д.т.н. И. Н. Сеницын
проф., д.ф.-м.н. В. Г. Ушаков
д.ф.-м.н. А. К. Горшенин — отв. секретарь
к.ф.-м.н. С. А. Христочевский

Редакция

к.ф.-м.н. Е. Н. Арутюнов
С. Н. Стригина

© Федеральный исследовательский центр «Информатика
и управление» Российской академии наук, 2023

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science

Журнал входит в систему Российского индекса научного цитирования (РИНЦ)
Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory
и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНИТИ
и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 33 № 3 Год 2023

СОДЕРЖАНИЕ

К 40-летию выхода постановлений ЦК КПСС и Совета Министров СССР о дальнейшем развитии работ в области вычислительной техники, в том числе в Академии наук СССР <i>В. Н. Захаров</i>	4
Анализ монотонного тренда временного ряда <i>М. П. Кривенко</i>	17
Диспетчеризация в частично наблюдаемых стохастических системах конечной емкости с параллельным обслуживанием <i>М. Г. Коновалов, Р. В. Разумчик</i>	29
Принципы разработки программного комплекса центра управления национальной исследовательской компьютерной сетью России <i>А. Г. Абрамов, В. А. Порхачёв</i>	48
Анализ эффективности алгоритма редукции в решении задачи об упаковке в контейнеры <i>Е. Б. Барашов, А. В. Егоркин, Д. В. Лемтюжникова, М. А. Посыпкин</i>	61
Методология научно-исследовательской работы при создании функций систем сканирования и печати <i>И. В. Сафонов, И. А. Матвеев</i>	76
Проблемы формирования системы согласованных локальных географических онтологий <i>Д. А. Никишин</i>	85
Некоторые подходы к анализу факторов, влияющих на информационную безопасность систем искусственного интеллекта <i>А. А. Зацаринный, А. П. Сучков</i>	95

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 33 № 3 Год 2023

СОДЕРЖАНИЕ

Некоторые проблемы мониторинга информационной безопасности критической инфраструктуры <i>А. А. Грушо, Н. А. Грушо, М. И. Забейло, А. А. Зацаринный, Е. Е. Тимонина</i>	108
Машинный перевод с помощью ChatGPT: мониторинг воспроизводимости результатов <i>А. Ю. Егорова, И. М. Зацман, В. О. Романенко</i>	117
Методический подход к контролю качества трансформации данных в междисциплинарной цифровой платформе <i>А. А. Зацаринный, А. П. Шабанов</i>	129
Теория S-символов: сетевой табс-решатель S-задач <i>В. Д. Ильин</i>	141
Очистка данных в технологии поддержки конкретно-исторических исследований <i>И. М. Адамович, О. И. Волков</i>	149
Социальная эффективность информационных технологий <i>К. К. Колин</i>	161
Об авторах	172
Правила подготовки рукописей статей	175
Requirements for manuscripts	179

К 40-ЛЕТИЮ ВЫХОДА ПОСТАНОВЛЕНИЙ ЦК КПСС И СОВЕТА МИНИСТРОВ СССР О ДАЛЬНЕЙШЕМ РАЗВИТИИ РАБОТ В ОБЛАСТИ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ, В ТОМ ЧИСЛЕ В АКАДЕМИИ НАУК СССР

*В. Н. Захаров*¹

Аннотация: В 2023 г. исполняется 40 лет с момента образования Института проблем информатики АН СССР, в настоящее время входящего в состав Федерального исследовательского центра «Информатика и управление» РАН. Институт был образован в соответствии с Постановлением ЦК КПСС и Совета Министров (СМ) СССР и Постановлением Совета Министров СССР от 29 июля 1983 г. В статье рассматривается некоторая предыстория развития вычислительной техники в СССР до выхода этих постановлений, возвращение тематики работ по вычислительной технике и информатике в АН СССР, создание Отделения информатики, вычислительной техники и автоматизации АН СССР, основные положения постановлений. Приведены также краткие сведения о развитии ИПИАН от момента создания до настоящего времени.

Ключевые слова: вычислительная техника; информатика; АН СССР; ЦК КПСС; Совет Министров СССР; Институт проблем информатики

DOI: 10.14357/08696527230301

EDN: XIWGAG

1 Введение

В СССР исследования в области создания электронных вычислительных машин (ЭВМ) начались в конце 1940-х гг. Первое в СССР авторское свидетельство на изобретение автоматической цифровой вычислительной машины (АЦВМ) на имя И. С. Брука и Б. И. Рамеева было датировано 4 декабря 1948 г. (приоритет) и выдано 16 февраля 1950 г. Создание первых в СССР ЭВМ — М1 (И. Брука) и МЭСМ (малая электронная счетная машина) (С. Лебедева) — датируется 1951 г. [1]. В 1950-е гг. активно начались работы в области разработки новых моделей ЭВМ и программного обеспечения для них. Первые отечественные машины создавались в академических институтах. В Москве работы велись под руководством члена-корреспондента АН СССР И. С. Брука сначала в Энергетическом институте АН СССР (ЭНИИ), где он организовал Лабораторию электросистем и начал исследования в области расчетов мощных энергетических систем. В 1956 г. на базе Лаборатории электросистем ЭНИИ АН СССР для разработки электронно-вычислительной техники была образо-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vzakharov@ipiran.ru

вана самостоятельная Лаборатория управляющих машин и систем (ЛУМС) АН СССР под руководством И. С. Брука. В 1957 г. он поставил научную проблему «Разработка теории, принципов построения и применения специализированных вычислительных и управляющих машин». Для ее решения 1 октября 1958 г. ЛУМС АН СССР была преобразована в Институт электронных управляющих машин (ИНЭУМ) АН СССР, директором которого стал И. С. Брук.

В Киеве под руководством академика АН УССР С. А. Лебедева работы по созданию МЭСМ велись в Лаборатории по спецмоделированию и вычислительной технике Института электротехники Академии наук Украинской ССР. В 1951 г. С. А. Лебедев переехал в Москву, где продолжил работу в Институте точной механики и вычислительной техники АН СССР (ИТМиВТ), который возглавлял с 1953 по 1973 гг.

Кроме упомянутых двух институтов в 1950-е гг. работы, связанные в основном с созданием средств программного обеспечения ЭВМ, велись в целом ряде академических институтов, таких как Институт прикладной математики, Вычислительный центр АН СССР, Институт автоматики и телемеханики АН СССР.

Работы по созданию аппаратных средств вычислительной техники и систем, базирующихся на этих средствах, проводились на предприятиях ряда промышленных министерств. Здесь можно упомянуть Специальное конструкторское бюро № 245 (СКБ-245), организованное в 1948 г., на базе которого в 1958 г. был образован НИИ электронных машин Минрадиопрома СССР. Научно-производственное предприятие (НПП) «Рубин» (г. Пенза) было образовано Постановлением СМ СССР от 23.02.1953 как филиал Московского СКБ-245. Постановлением ЦК КПСС и СМ СССР от 06.10.1958 филиал СКБ-245 был преобразован в НИИ управляющих вычислительных машин (НИИУВМ), а через 8 лет Приказом Минрадиопрома СССР от 24.03.1966 переименован в Пензенский НИИ математических машин. В Москве работали завод счетных аналитических машин (САМ) и НИИСчетмаш. В 1957 г. Загорский электромеханический завод «Звезда» Минрадиопрома СССР был определен соразработчиком ИТМиВТ. В 1951 г. на правительственном уровне было принято решение о строительстве Казанского завода математических машин (КЗММ), который вошел в число первых четырех заводов СССР по выпуску ЭВМ. С 1963 г. завод был подчинен Минрадиопрому СССР и назван Казанским заводом электронных вычислительных машин (КЗЭВМ). В 1958 г. при Минском заводе математических машин Минрадиопрома СССР было создано Специальное конструкторское бюро (впоследствии — НИИЭВМ). В Ленинграде в 1956 г. была создана Специальная лаборатория (СЛ-11) в рамках ОКБ-998 Авиапрома, позднее превратившаяся в Конструкторско-технологическое бюро «Светлана-микроэлектроника» (КБСМ). В начале августа 1954 г. был создан Вычислительный центр № 1 Министерства обороны, который спустя 7 лет был преобразован в 27 ЦНИИ МО СССР, сыгравший значимую роль в создании систем на базе средств вычислительной техники, в первую очередь в военном применении.

В начале 1960-х гг. в рамках так называемых «хрущевских реформ» руководство страны приняло решение, направленное на «приближение науки к производству». Значительная часть институтов бывшего Отделения технических наук АН СССР была передана в промышленность, среди которых были ведущие институты в области вычислительной техники (ИНЭУМ, ИТМиВТ). Постепенно была изменена тематика этих институтов в ущерб фундаментальным исследованиям и перспективным разработкам. Изменилась и мотивация ученых, работавших в этих, теперь уже промышленных, институтах. Это решение привело к значительному ослаблению позиций АН СССР по направлениям науки, связанным с кибернетикой и информатикой. Оно коснулось прежде всего Российской Федерации, академическая наука которой отождествлялась с АН СССР. Работы в области кибернетики и информатики стали развиваться в это время в Академиях наук союзных республик (Украины, Белоруссии, Эстонии, Латвии, Литвы, Армении, Грузии, Узбекистана), а также в Сибирском и Дальневосточном отделениях АН СССР, благодаря усилиям их организаторов — академиков М. А. Лаврентьева, С. Л. Соболева, А. А. Воронова [2].

В конце 1967 г. было принято Постановление ЦК КПСС и Совета Министров СССР № 1180-420 о развитии средств производства вычислительной техники в стране, которым Минрадиопрому разрешалось создать Научно-исследовательский центр электронной вычислительной техники (НИЦЭВТ), который стал головным по созданию семейства ЭВМ «Ряд», повторявших архитектуру машин ИВМ.

2 Возвращение исследований в области информатики и вычислительной техники в Академию наук

Как уже упоминалось выше, начало 1960-х гг. было периодом, когда многие академические институты переводились в промышленные министерства. Не избежал этого и ИНЭУМ АН СССР — он был передан в Госэкономсовет СССР, который в дальнейшем стал Министерством приборостроения, средств автоматизации и автоматизированных систем управления (Минприбор СССР).

К началу 1980-х гг. в мире начался бум использования средств вычислительной техники буквально во всех областях. Это было вызвано в значительной степени появлением на рынке нового вида массовой техники — персональных ЭВМ. В СССР к этому времени в ряде министерств и ведомств велись работы по разработке и производству средств вычислительной техники, однако уже отчетливо проявилось осознание заметного отставания от мировых лидеров в этом направлении. В конце 1982 г. после смерти руководителя страны Л. И. Брежнев и избрания на пост генерального секретаря ЦК КПСС Ю. В. Андропова в стране начались определенные изменения, затронувшие и область вычислительной техники. Обновленным руководством страны был принят целый ряд

важных решений, направленных на преодоление этого отставания. По правилам того времени такие решения оформлялись в виде постановлений ЦК КПСС или Совета Министров СССР или в виде совместных постановлений этих двух органов [3].

В 1983 г. Общее собрание Академии наук СССР проходило 2 и 3 марта. В своем вступительном слове Президент АН СССР А. П. Александров говорил о необходимости возрождения и усиления фундаментальных исследований в области компьютерных наук, а также целесообразности (для обеспечения технологического паритета в области информационных технологий и вычислительной техники) создания специального отделения. С большим докладом «Об организации работ по информатике, вычислительной технике и автоматизации в Академии наук СССР» выступил вице-президент АН СССР Е. П. Велихов. В результате Постановлением Общего собрания АН СССР № 12 от 3 марта 1983 г. было принято решение об организации Отделения информатики, вычислительной техники и автоматизации (ОИВТА) в составе Секции физико-технических и математических наук при Президиуме АН СССР. Этим же постановлением указанной секции было поручено представить на утверждение очередного Общего собрания АН СССР персональный состав Отделения.

Распоряжением Президиума АН СССР от 24 ноября 1983 г. был утвержден состав организационного бюро ОИВТА под председательством академика Е. П. Велихова. В него вошли: академики Ж. И. Алферов, О. М. Белоцерковский, В. А. Мельников, члены-корреспонденты АН СССР К. А. Валиев, А. П. Ершов, Ч. В. Копецкий, Б. Н. Наумов, д. т. н. В. М. Пономарев и к. ф.-м. н. ученый секретарь Ю. С. Вишняков.

Постановлением Президиума АН СССР от 23 февраля 1984 г. были определены научные учреждения АН СССР, вошедшие в состав ОИВТА. Это были четыре ранее существовавших института: Институт прикладной математики им. М. В. Келдыша, Вычислительный центр, Институт проблем передачи информации и Ленинградский научно-исследовательский вычислительный центр, а также пять новых институтов, созданных постановлением ЦК КПСС и СМ СССР, речь о котором пойдет ниже.

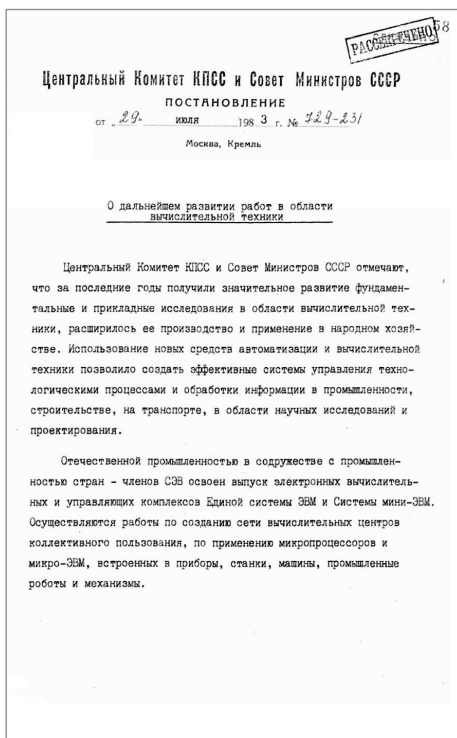
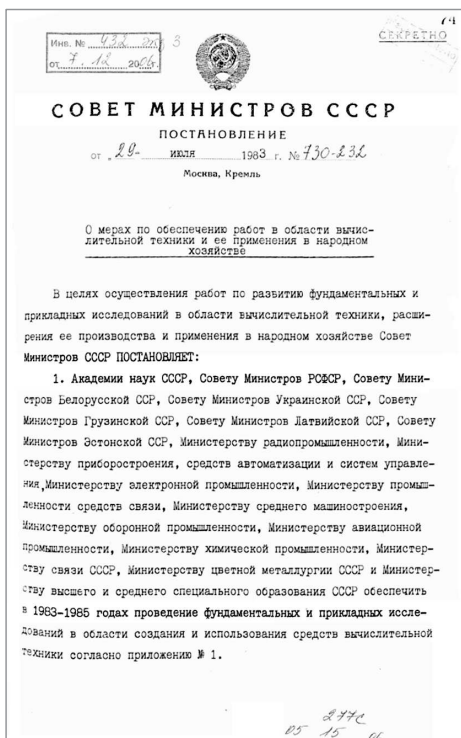
Общее собрание АН СССР, состоявшееся 14 марта 1984 г., своим постановлением № 10 утвердило персональный состав ОИВТА, в который вошли 10 действительных членов АН СССР (Белоцерковский О. М., Бункин Б. В., Велихов Е. П., Воронов А. А., Дородницын А. А., Мельников В. А., Пугачев В. С., Самарский А. А., Семенихин В. С. и Тихонов А. Н.) и 25 членов-корреспондентов (Алексеев А. С., Бицадзе А. В., Бурцев В. С., Валиев К. А., Говорун Н. Н., Гуляев Ю. В., Евтихийев Н. Н., Емельянов С. В., Ершов А. П., Золотов Е. В., Копецкий Ч. В., Королев Л. Н., Лавров С. С., Лопато Г. П., Макаров И. М., Моисеев Н. Н., Наумов Б. Н., Попов Е. П., Пospelов Г. С., Ржанов А. В., Савин А. И., Сифоров В. И., Тихомиров В. В., Цыпкин Я. З. и Шереметьевский Н. Н.).

3 Постановления ЦК КПСС и Совета Министров СССР от 29 июля 1983 года

В этом году исполняется 40 лет выпущенным в один день 29 июля 1983 г. совместному Постановлению ЦК КПСС и Совета Министров СССР № 729-231 «О дальнейшем развитии работ в области вычислительной техники» [4] и детализирующему его Постановлению Совета Министров СССР № 730-232 «О мерах по обеспечению работ в области вычислительной техники и ее применения в народном хозяйстве» [5], сыгравших большую роль в истории информатики в стране (см. рисунок).

В постановлении ЦК КПСС и СМ СССР, под которым стоят подписи секретаря ЦК КПСС Ю. Андропова и Председателя СМ СССР Н. Тихонова, были сформулированы главные задачи в области исследований, разработки, производства и применения вычислительной техники в народном хозяйстве на ближайшую перспективу:

- проведение фундаментальных и прикладных исследований по созданию средств вычислительной техники на основе перспективной конструктивно-



Копии первых страниц постановлений ЦК КПСС и СМ СССР

- элементной базы, создание новых высокопроизводительных вычислительных и управляющих комплексов (в том числе специализированных) и сетей центров общего пользования, расширение областей эффективного применения средств вычислительной техники;
- разработка и освоение производства технических, а также программных средств вычислительной техники и носителей информации, соответствующих по техническому уровню лучшим мировым образцам;
 - повышение надежности и совершенствование структуры выпуска средств вычислительной техники, унификация и стандартизация технических, конструктивных и программных решений, улучшение обеспечения вычислительных и управляющих комплексов периферийными устройствами и поставка этих комплексов в конфигурациях, соответствующих требованиям пользователей вычислительной техники;
 - дальнейшее развитие и интеграция автоматизированных систем различного уровня и назначения, а также повышение экономической эффективности их в народном хозяйстве;
 - создание систем передачи и телеобработки данных для сетей вычислительных центров коллективного пользования (ВЦКП);
 - широкое применение микропроцессоров и микро-ЭВМ при одновременном снижении их стоимости;
 - организация разработки и производства на промышленной основе программных средств для вычислительных и управляющих комплексов и автоматизированных систем;
 - подготовка специалистов по разработке, обслуживанию и применению технических и программных средств вычислительной техники;
 - расширение круга пользователей вычислительной техникой путем организации системы переподготовки специалистов во всех отраслях народного хозяйства, включая руководящих работников.

На Академию наук СССР этим постановлением была возложена ответственность за проведение фундаментальных и прикладных исследований в области вычислительной техники и автоматизированных систем, а также научное руководство такими работами, проводимыми организациями Академий наук союзных республик, высшей школы, министерств и ведомств. Было предписано считать это одной из важнейших задач Академии наук СССР.

Именно в этом совместном постановлении ЦК КПСС и СМ СССР было сказано: «Принять предложение АН СССР, согласованное с ГКНТ и комиссией Президиума СМ СССР по военно-промышленным вопросам о создании в системе АН СССР: Научного центра по фундаментальным проблемам вычислительной техники и систем управления (с включением в его состав организуемых в г. Ярославле Института проблем вычислительной техники, Института микроэлектроники, СКБ и опытного производства) на правах научно-технического

объединения; Института проблем кибернетики на базе лабораторий научного совета АН СССР по комплексной проблеме «Кибернетика» в г. Москве с филиалом в г. Переславль-Залесский Ярославской области; Института проблем технологии микроэлектроники и особо чистых материалов (со специальным конструкторско-технологическим бюро и опытным производством) в пос. Черноголовка на базе ряда подразделений Института физики твердого тела и других организаций АН СССР. Разрешить АН СССР создать Институт проблем информатики с опытным производством в г. Москве и с филиалами в гг. Казани и Бердянске Запорожской области». Эти созданные организации в 1984 г. вошли в состав ОИВТА.

В совместном постановлении были утверждены «Основные направления фундаментальных и прикладных исследований в области вычислительной техники и автоматизированных систем на период до 1990 года». Приведем их здесь полностью для того, чтобы почувствовать «аромат эпохи».

ЭВМ и вычислительные системы

Научная разработка единой стратегии развития вычислительных и автоматизированных управляющих систем на основе унификации и стандартизации технических и программных средств.

Создание ЭВМ сверхвысокой производительности (свыше 1 млрд операций в секунду) с использованием новых принципов организации вычислительного процесса и построения архитектуры ЭВМ.

Разработка новых принципов создания специализированных микропроцессоров, микро- и мини-ЭВМ.

Разработка перспективных оперативных и внешних запоминающих устройств (ЗУ) и широкого набора периферийных средств.

Исследование и разработка перспективных технических средств вычислительной техники с использованием новых физических принципов (криоэлектроники, оптоэлектроники, акусто- и магнитоэлектроники).

Разработка языков формализованного описания архитектуры ЭВМ для систем автоматизированного проектирования машин.

Исследования в области искусственного интеллекта.

Программное обеспечение

Исследование и создание методов математического моделирования разрабатываемых ЭВМ.

Создание современной технологии разработки программного обеспечения, включая разработку методов автоматизированного построения программных систем.

Исследование и разработка машиннезависимых операционных систем.

Исследование и разработка языков программирования высокого уровня, в том числе логических (непроцедурных), метаязыков.

Микроэлектронная база

Исследование физических и технологических проблем создания субмикронных структур на перспективных материалах, создание опытных образцов технологического оборудования.

Разработка новых принципов создания сверхбыстродействующих и сверхбольших интегральных схем и их макетирование.

Разработка новых методов контроля полупроводниковых материалов и структур сверхбыстродействующих и сверхбольших интегральных схем.

Разработка технологии изготовления специальных и особо чистых материалов и совершенных кристаллов.

Разработка методов и средств контроля технологических процессов и качества изделий микроэлектронной базы.

Изучение факторов, определяющих долговременную стабильность элементной базы микроэлектроники.

Создание банков данных для разработки и оптимизации процессов изготовления особо чистых материалов, сверхбыстродействующих и сверхбольших интегральных схем.

Применение вычислительной техники

Исследование принципов и основных направлений создания сетей ЭВМ и ВЦКП.

Разработка методологии создания комплексных интегрированных автоматизированных систем, охватывающей все этапы разработки и изготовления этих систем.

Разработка теоретических основ и методологии построения взаимоувязанных автоматизированных систем различных уровней и направлений.

В Постановлении СМ СССР [5], подписанном Председателем СМ СССР Н. Тихоновым и управляющим делами СМ СССР М. Смиртюковым, были определены конкретные задачи по проведению фундаментальных и прикладных исследований, перечислены задания на разработку технических заданий, изготовление опытных образцов и освоение серийного производства новых средств вычислительной техники на несколько ближайших лет. Этим же постановлением предусматривался целый комплекс мер, необходимых для реализации поставленных задач. Так, было предписано осуществить строительство, расширение и реконструкцию предприятий и организаций по списку с выделением средств, разработку программы стандартизации и межотраслевой унификации средств вычислительной техники, создание хозрасчетных вычислительных центров. Были и такие непривычно сейчас звучащие слова: «ГКНТ, Комиссии по ВПК и АН СССР осуществить перевод в 6-месячный срок до 300 квалифицированных специалистов из МРП, Минприбора и МЭП в ИПИ АН. Разрешить АН СССР разместить ИПИ АН в здании Президиума АН СССР, строительство которого предусмотрено распоряжением СМ СССР от 01.12.1975 № 2614. Увеличить на

1983 г. АН СССР лимит численности работников организаций, расположенных в г. Москве, на 100 человек».

С исторической точки зрения представляют интерес задания по созданию средств вычислительной техники, содержащие конкретные параметры, характеризующие уровень того времени и ближайшие перспективы. Приведем лишь некоторые примеры.

Так, ставилась задача по исследованию принципов создания накопителей на магнитных дисках (НМД) емкостью до 3000 МБайт, со скоростью обмена данными до 3000 КБайт/с, средним временем доступа 20–25 мс; по разработке методов создания ЗУ на цилиндрических магнитных доменах (ЦМД) емкостью 10^8 бит на кристалл; по исследованию принципов создания оптико-механического дискового ЗУ емкостью $(3-5) \cdot 10^{12}$ бит, со скоростью обмена данными до 2 МБайт/с, средним временем доступа 15 с. Ставилась задача по производству в 1985 г. накопителей на гибких магнитных дисках емкостью 0,14 МБайт, скоростью обмена 250 кбит/с, с плотностью записи 109/218 бит/мм, диаметром диска 130 мм (НИИ периферийного оборудования, Киев). В части мейнфреймов в задании на 1984 г. фигурировала ЭВМ ЕС-1046 производительностью 1 млн оп/с, емкостью ОЗУ до 8 МБайт (разработка и изготовление опытного образца НИЦЭВТ и ЕРНИИММ, серийное производство — Казанский завод ЭВМ МРП). А на 1988 г. было задано уже производство ЭВМ ЕС-1087 производительностью 18 млн оп/с, емкостью ОЗУ до 128 МБайт (Минское производственное объединение вычислительной техники, МПО ВТ). В части мини- и микроЭВМ говорилось о запуске в производство в 1984 г. вычислительного комплекса СМ-1420, программно совместимого с ЭВМ СМ-3 и СМ-4, разрядностью 16 двоичных разрядов, производительностью до 1 млн оп/с, емкостью ОЗУ 2 МБайт (ИНЭУМ и Киевское ПО «Электронмаш» им. В. И. Ленина Минприбора), и микроЭВМ «Электроника-60-1» разрядностью 16 двоичных разрядов, производительностью до 600 тыс. оп/с, емкостью ОЗУ до 256 КБайт (ПО «Электроника» Минэлектронпрома).

О масштабах проводимых работ можно судить, например, по перечню, содержащемуся в задании по строительству новых, расширению и реконструкции действующих предприятий и организаций, задействованных в выполнении постановления. Приведем его полностью.

Минрадиопром: НИЦЭВТ (Москва), Загорский электромеханический завод (Московская область), Костромской электромеханический завод, завод САМ (Москва), Волжский завод радиотехнических элементов (Волгоградская область), Астраханский машиностроительный завод «Прогресс», Пензенский завод «ВЭМ», Минский завод ЭВМ, Минский завод многослойных печатных плат, Кишиневский завод счетных машин, Тбилисский завод электронно-вычислительной техники, Ташкентский завод электронных вычислительных машин «Алгоритм», Кировский приборостроительный завод, НИИ ЭВМ (Минск), Казанский завод пишущих устройств, Казанский завод ЭВМ, Боярский машиностроительный завод «Искра» (Киевская область), Брестский электромеханический завод,

завод «Электроприбор» (г. Каменец-Подольский Хмельницкой области), Каневский электромеханический завод «Магнит» (Черкасская область), Фрунзенский завод ЭВМ, НИИ вычислительной техники (Пенза), Бакинский завод ЭВМ, Бакинский радиозавод, Казанский НИТИ вычислительной техники, КБ «Север» (Киров), ПО «ЦентрЭВМкомплекс» (Москва).

Минприбор: Орловский завод УВМ, Киевский завод электронных вычислительных и управляющих машин, Северодонецкий приборостроительный завод, Вильнюсский завод счетных машин, Рязанский завод САМ, Таурагский завод элементов вычислительных машин (Литовская ССР), Паневежский завод точной механики (Литовская ССР), Курский завод «Счетмаш».

Минэлектронпром: завод «Альтаир» (Ярославль), завод «Процессор» (Воронеж), ПО «Электроника» (Воронеж), Брянский завод полупроводниковых приборов, Кишиневский завод «Мезон», НИИ физических проблем (Зеленоград), НИИ молекулярной электроники (Зеленоград), ЦНИИ «Циклон» (Москва), завод микропроцессоров (Бельцы Кишиневской области), завод «Девиз» (г. Алексеевка Белгородской области), Калининградский машиностроительный завод, Выборгский приборостроительный завод, Черняховский машиностроительный завод (Калининградская область).

АН СССР: Научный центр АН СССР (Ярославль), филиал Института кибернетики АН СССР (Переславль-Залесский), Институт проблем кибернетики АН СССР, лаборатория микроэлектроники Института общей физики АН СССР и Всесоюзный научно-исследовательский центр по изучению свойств поверхности и вакуума Госстандарта и АН СССР (Москва), Институт проблем технологии микроэлектроники и особо чистых материалов с СКБ (Черноголовка), Институт физики твердого тела (Черноголовка), Институт проблем информатики (15 тыс. кв. м производственной площади и 5 тыс. кв. м жилой), Саратовский филиал ИРЭ.

4 40-летие Института проблем информатики

Постановление ЦК КПСС и Совета Министров СССР, разрешающее АН СССР создать Институт проблем информатики АН СССР (ИПИАН), было принято 29 июля 1983 г. [6], а уже 2 августа 1983 г. вышло соответствующее распоряжение Президиума АН СССР.

В начале 1980-х гг. появляются и стремительно завоевывают новые области применения массовые средства вычислительной техники — персональные компьютеры, существенное развитие получают научные основы информатики. Однако «ведомственность» организаций, занимающихся проблемами информатизации, стала тормозом для развития информатики в СССР. Именно с целью преодоления «ведомственных барьеров», а также развития методов и средств информатизации было образовано новое отделение АН СССР, а член-корреспондент АН СССР Б. Н. Наумов, возглавлявший Институт электронных

управляющих машин Минприбора СССР, был назначен директором-организатором одного из новых институтов — ИПИАН. Основная задача ИПИАН была определена как «проведение фундаментальных и прикладных исследований в области технических и программных средств массовой вычислительной техники и систем на их основе», а интеллектуальной базой первых научных подразделений ИПИАН стали коллективы ряда научных отделов ИНЭУМ, переведенные в ИПИАН в начале 1984 г.

Институт в первые годы своего существования быстро развивался — в его состав вошли филиалы в Бердянске, Казани, Орле. В 1990 г. был образован совместный отдел с радиотехническим институтом в г. Таганроге. Общая численность сотрудников института в 1988–1989 гг. превышала 1000 человек. В 1989 г. директором института стал член-корреспондент АН СССР Игорь Александрович Мизин (в 1997 г. избран действительным членом РАН). В 1999 г. директором института стал Игорь Анатольевич Соколов (в 2003 г. избран членом-корреспондентом РАН, а в 2008 г. — действительным членом РАН). В 1992 г. институт получил новое наименование — Институт проблем информатики Российской академии наук (ИПИ РАН). В настоящее время преемником Института проблем информатики РАН стал Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) — подведомственная Министерству науки и высшего образования Российской Федерации научная организация, выполняющая фундаментальные, поисковые и прикладные научные исследования и разработки в области вычислительной и прикладной математики, системного анализа и управления, теоретической информатики и информационных технологий, развития информационно-телекоммуникационной инфраструктуры и информатизации общества. ФИЦ ИУ РАН был создан в декабре 2014 г. путем реорганизации Федерального государственного бюджетного учреждения науки Института проблем информатики Российской академии наук (ИПИ РАН) в форме присоединения к нему Федерального государственного бюджетного учреждения науки Института системного анализа Российской академии наук (ИСА РАН) и Федерального государственного бюджетного учреждения науки Вычислительного центра им. А. А. Дородницына Российской академии наук (ВЦ РАН). Центр имеет филиалы в г. Орле и в г. Калининграде, которые до реорганизации были филиалами ИПИ РАН. Научно-методическое руководство ФИЦ ИУ РАН осуществляется двумя отделениями РАН — Отделением нанотехнологий и информационных технологий (до 2002 г. — Отделение информатики, вычислительной техники и автоматизации, в 2002–2007 гг. — Отделение информационных технологий и вычислительных систем) и Отделением математических наук.

Литература

1. *Zakharov V. N.* Two fates in the history of computer technology in the USSR (Lebedev and Brouk) // 4th Conference (International) on Computer Technology in Russia and in

the Former Soviet Union Proceedings / Eds. I. Krayneva, A. Tomilin. — Piscataway, NJ, USA: IEEE, 2018. P. 1–4. doi: 10.1109/SoRuCom.2017.00006.

2. Отечественная электронная вычислительная техника / Сост. С. А. Муравьев; под ред. С. В. Хохлова. — М.: Столичная энциклопедия, 2014. 400 с.
3. *Zakharov V.* Computers and their application in the USSR in the middle of the 1980s: Situation, actions taken, predictions of development // 3rd Conference (International) on Computer Technology in Russia and in the Former Soviet Union Proceedings. — Piscataway, NJ, USA: IEEE, 2014. P. 53–60. doi: 10.1109/SoRuCom.2014.19.
4. О дальнейшем развитии работ в области вычислительной техники: Постановление ЦК КПСС и СМ СССР от 29.07.1983 № 729-231.
5. О мерах по обеспечению работ в области вычислительной техники и ее применения в народном хозяйстве: Постановление СМ СССР от 29.07.1983 № 730-232.
6. *Захаров В. Н.* Работы Института проблем информатики РАН // История отечественной электронной вычислительной техники / Под ред. А. С. Якунина. — М.: Столичная энциклопедия, 2014. С. 444–448.

Поступила в редакцию 31.07.23

**TO THE 40TH ANNIVERSARY OF THE DECREES
OF THE CENTRAL COMMITTEE OF THE CPSU
AND THE COUNCIL OF MINISTERS OF THE USSR
ON THE FURTHER DEVELOPMENT OF WORK
IN THE FIELD OF COMPUTER TECHNOLOGY,
INCLUDING AT THE USSR ACADEMY OF SCIENCES**

V. N. Zakharov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The year 2023 marks the 40th anniversary of the establishment of the Institute of Problems of Informatics of the USSR Academy of Sciences, currently part of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences. The Institute was established in accordance with a decree of the Central Committee of the CPSU and the Council of Ministers of the USSR and a decree of the Council of Ministers of the USSR dated July 29, 1983. The article discusses some prehistory of the development of computer technology in the USSR before these decrees were issued, the return of the subject of work on computer technology and computer science to the USSR Academy of Sciences, the establishment of the Department of Computer Science, Computer Technology, and Automation of the USSR Academy of Sciences, and the main provisions of the decrees. Brief information about the development of IPIAN from the moment of its creation to the present time is also provided.

Keywords: computer engineering; computer science; Academy of Sciences of the USSR; Central Committee of the CPSU; Council of Ministers of the USSR; Institute of Informatics Problems

DOI: 10.14357/08696527230301

EDN: XIWGAG

References

1. Zakharov, V. N. 2018. Two fates in the history of computer technology in the USSR (Lebedev and Brouk). *4th Conference (International) on Computer Technology in Russia and in the Former Soviet Union Proceedings*. Eds. I. Krayneva and A. Tomilin. Piscataway, NJ: IEEE. 1–4. doi: 10.1109/SoRuCom.2017.00006.
2. Murav'ev, S. A., compiler. 2014. *Otechestvennaya elektronnyaya vychislitel'naya tekhnika: Biograficheskaya entsiklopediya* [Domestic electronic computer technology: Biographical encyclopedia]. Ed. S. V. Khokhlov. Moscow: Capital Encyclopedia. 400 p.
3. Zakharov, V. N. 2014. Computers and their application in the USSR in the middle of the 1980s: Situation, actions taken, predictions of development. *3rd Conference (International) on Computer Technology in Russia and in the Former Soviet Union Proceedings*. Piscataway, NJ: IEEE. 53–60. doi: 10.1109/SoRuCom.2014.19.
4. O dal'neyshem razvitii rabot v oblasti vychislitel'noy tekhniki: Postanovlenie TsK KPSS i SM SSSR ot 29.07.1983 No. 729-231 [On the prospects for the development of work in the field of computer technology: Decree of the Central Committee of the CPSU and the Council of Ministers of the USSR No. 729-231 dated 29.07.1983].
5. O merakh po obespecheniyu rabot v oblasti vychislitel'noy tekhniki i ee primeneniya v narodnom khozyaystve: Postanovlenie SM SSSR ot 29.07.1983 No. 730-232 [On measures to ensure work in the field of computer technology and its application in the national economy: Decree of the Council of Ministers of the USSR No. 730-232 dated 29.07.1983].
6. Zakharov, V. N. 2014. Raboty Instituta problem informatiki RAN [Proceedings of the Institute of Informatics Problems of the Russian Academy of Sciences]. *Istoriya otechestvennoy elektronnoy vychislitel'noy tekhniki* [History of domestic computing technology]. Ed. A. S. Yakunin. Moscow: Capital Encyclopedia. 444–448.

Received July 31, 2023

Contributor

Zakharov Victor N. (b. 1948) — Doctor of Science in technology, associate professor, scientific secretary, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vzakharov@ipiran.ru

АНАЛИЗ МОНОТОННОГО ТРЕНДА ВРЕМЕННОГО РЯДА

*М. П. Кривенко*¹

Аннотация: Рассматривается задача выявления изменений за период наблюдения характеристик исследуемого временного ряда. Практически всегда тренд возникает на фоне статистической зависимости элементов временного ряда. Наличие зависимости отдельных наблюдений становится мешающим фактором. При построении предположений об изменениях предпочтение отдается моделям общего характера: тренд вероятностных характеристик есть монотонная функция неизвестного вида от номера наблюдения (монотонный тренд). Необходимость учета в модели временного ряда объединения обоих действующих факторов и потребность получать работоспособные методы обработки данных приводят к следующей схеме действий: принять за основу уже разработанные процедуры, далее по возможности подогнать условия их корректного применения под требуемые и, наконец, использовать приспособленные варианты. Проводится анализ методов обработки мешающего фактора: нейтрализации зависимости, учета зависимости, обобщения модели временного ряда. В качестве примера рассматривается задача контроля стабильной работы двухпроцессорной системы обработки заданий со случайным выбором числа требуемых процессоров. Демонстрируются возможности и ограничения предложенных методов.

Ключевые слова: монотонный тренд; декорреляция; ARIMA-модели

DOI: 10.14357/08696527230302

EDN: WVGUVX

1 Введение

Рассматривается задача выявления изменений характеристик исследуемого временного ряда. Его неотъемлемая особенность заключается в том, что отдельные наблюдения могут быть зависимыми. Природа зависимости между элементами ряда иногда известна и даже достаточно просто моделируется (например, выходной поток заявок в системе массового обслуживания с известной структурой), а иногда ставит в тупик исследователей (например, результаты повторно проводимых медицинских анализов признанного здоровым пациента); в данной статье наличие зависимости считается мешающим фактором. Изменения же характеристик наблюдаемого процесса представляют первейший интерес, при построении соответствующих предположений предпочтение по возможности отдается моделям общего характера. Причины этого состоят в том, что подчас нет достаточно полной информации для построения простых адекватных практике теоретико-вероятностных схем либо в априорные предположения изначально

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mkrivenko@ipiran.ru

заложен достаточный произвол. Описанная постановка задачи востребована, например, при лонгитюдном анализе данных в медицине, для сопровождения социологических процессов, в случае отслеживания экстремальных климатических явлений, при управлении доходами в эконометрике и т. д. Интересно обратить внимание еще на одну прикладную область анализа данных — это статистический контроль процессов, систем, устройств и т. п., в частности мониторинг функционирования систем обслуживания с очередями (см., например, [1]). Эта область по-прежнему востребована, по отдельным задачам достаточно нова и активно формируется в настоящее время. Кроме того, объектом контроля часто выступают технические средства, функционирование которых можно моделировать, и, следовательно, становится возможным хотя бы частично снять напряженность, связанную с обычно бедным по объему статистическим материалом.

Далее в основу модели данных положено предположение о том, что изменение их вероятностных характеристик есть монотонная функция неизвестного вида от номера наблюдения (далее — монотонный тренд). Кратко теоретические основы статистического анализа данных для этой модели изложены в [2]. К ее основным необходимым для данной работы моментам необходимо отнести следующие.

Случай тренда среднего распределения Гаусса для последовательности независимых наблюдений наиболее разработан. Для получения оценки тренда и принятия решения о его наличии используется PAV-процедура (Pool-Adjacent-Violators) устранения неверных неравенств. Ее основной шаг заключается в объединении соседних групп наблюдений, которые не дают удовлетворяющие требуемым условиям обычные оценки среднего. Проверка гипотезы об однородности выборки против конкурирующей гипотезы о монотонной упорядоченности осуществляется с помощью критерия отношения правдоподобия. Их статистики в зависимости от априорных предположений принимают ту или иную форму, важным элементом которой служит так называемая ступенька — компоненты вектора оценки среднего с одинаковыми значениями. Основным результатом относительно поведения этих статистик при нулевой гипотезе заключается в том, что их условные по числу ступенек распределения суть известные χ^2 -распределения и β -распределение. В результате центральной становится задача нахождения $\tau_N(j)$ — вероятности того, что оценка максимального правдоподобия монотонного тренда, полученная по N наблюдениям, содержит j ступенек. Было доказано, что

$$\tau_N(j) = \frac{|s(N, j)|}{N!}, \quad j = 1, \dots, N,$$

где $s(N, j)$ — числа Стирлинга 1-го рода. Позже было обосновано, что при этом необязательно рассматривать выборочное среднее [3]. Это говорит в пользу устойчивости статистического вывода на основе критерия числа ступенек.

Переход в общем случае к модели непараметрического тренда целесообразно осуществлять так, чтобы расширить область применения уже построенных ре-

шений. Здесь наиболее практичным оказывается использование свободных от распределения процедур, заключающееся в переходе от наблюдений к их рангам, и рассмотрение в качестве модели тренда стохастической упорядоченности.

2 Методы обработки зависимости элементов временного ряда

Практически всегда тренд возникает на фоне статистической зависимости элементов временного ряда. Попытка объединения этих двух факторов приводит к популярной модели (например, в задачах описания глобальной температуры при климатических исследованиях) анализа зависимости от времени в виде тренда определенного вида и процесса авторегрессии с нулевым средним. Хотя были предприняты значительные усилия для разработки тестов выявления, в частности линейного тренда при наличии последовательной корреляции, процедуры тестирования, используемые на практике, нельзя признать в общем случае удовлетворительными (см., например, [4]). При этом априорные предположения относительно анализируемых данных оказываются настолько жесткими, что становятся непрактичными. Поэтому необходимость учета в модели временного ряда объединения всех действующих факторов (т. е. задание распределения элементов временного ряда в совокупности) и потребность получать работоспособные методы обработки данных (в особенности на разведочном этапе исследований) приводят к следующей схеме действий: принять за основу уже разработанные процедуры, далее по возможности подогнать условия их корректного применения под требуемые и, наконец, использовать приспособленные варианты.

Нейтрализация зависимости. Если позволяют условия задачи анализа данных, то радикальным приемом может стать ослабление влияния или в идеале устранение зависимости, но так, чтобы наличие тренда осталось без изменений. Для этого используется термин «декорреляция» — это любой процесс, который приводит к уменьшению автокорреляции во временном ряде при сохранении других его аспектов. Практика анализа данных (см., например, [5]) подсказывает целесообразность использования для измерения эффективности декорреляции временной сдвиг (лаг), для которого автокорреляционная функция достигает заданного значения. Далее в качестве такого значения берется нулевое (в выборочном случае в виде доверительного интервала с надежностью в 1%), а соответствующий лаг минус 1 условно назовем «глубиной» зависимости.

Применительно к задачам анализа тренда декорреляция — это в первую очередь прореживание исходных данных. Оно может быть реализовано, во-первых, безотносительно к механизму порождения временного ряда или, во-вторых, с его учетом. При первом подходе естественно обратиться к случайному выбору или же оставлять для анализа только элементы временного ряда, отстоящие друг от друга с определенным лагом. Вторая идея зависит от источника формирования данных, и ее реализация приводит к комбинированной процедуре: вначале выделение серии последовательных элементов временного ряда, в которой «внутренние» элементы согласно механизму порождения временного ряда приводят

к взаимосвязи «крайних» элементов, что порождает прореженный временной ряд «крайних» элементов, к которому дополнительно применяется случайный или детерминированный выбор.

В качестве иллюстрации предлагаемых решений рассмотрим задачу контроля стабильной работы двухпроцессорной системы обработки заданий со случайным выбором числа требуемых процессоров [6]. Функционирование подобной системы $M/M/2$ определяется параметрами: λ — интенсивность поступления заданий на обработку; p_1 — вероятность того, что для выполнения задания требуется один процессор; μ — среднее время обработки задания процессором/процессорами. Согласно ссылке в разд. 3 работы [6] условие стабильности функционирования этой системы принимает вид:

$$\frac{\lambda}{\mu} < \frac{2}{2 - p_1^2}. \quad (1)$$

В качестве объекта статистического анализа выступает последовательность значений времени пребывания отдельных заданий в системе. Обработка данных проводится по фрагментам, длина которых выбрана стандартной и равной 10^4 , что не противоречит практике контроля функционирования систем обработки и позволяет провести сравнительный анализ ряда методов в совокупности.

Эффективное прореживание путем случайного выбора реализует переход от коррелированной последовательности к некоррелированной подпоследовательности максимальной длины, что достигается методом бисекции. Подбор шага детерминированного прореживания осуществляется последовательно от 1 до значения длины исходного временного ряда и завершается при встрече случая некоррелированности или при достижении длины исходной последовательности.

В основу комбинированной процедуры положено прореживание изолированием: для каждого фиксированного задания можно определить сопутствующую группу заданий, находящихся одновременно с ним в обработке, и оставить только те задания, для которых сопутствующие группы не пересекаются. Заметим, что такое сжатие данных не гарантирует отсутствия коррелированности в выходной последовательности, поэтому его приходится дополнять другими процедурами.

Для системы с параметрами $\lambda = 0,8$, $\mu = 1$ и $p_1 = 0$ были проведены 100 экспериментов, в ходе которых проводился сравнительный анализ эффективности (с точки зрения потери данных) трех процедур прореживания: случайной, детерминированной и комбинированной. Также проверялась гипотеза о нормальном распределении элементов временного ряда после прореживания. Для этого использовался сводный тест на основе коэффициентов асимметрии и эксцесса в редакции [7]. Если средняя «глубина» зависимости исходной последовательности данных составила 63, то средние значения доли оставшихся данных оказались равными 2% для случайного прореживания, 4% для детерминированного и 9% для комбинированного.

Полученные результаты позволяют сформулировать следующие выводы:

- детерминированная декорреляция приводит к меньшим потерям данных по сравнению со случайной;
- с учетом предыдущего пункта комбинированный метод, включающий прореживание изолированием с последующей детерминированной процедурой декорреляции, обладает наименьшими потерями данных из трех рассматриваемых преобразований;
- декорреляция позволяет корректно проводить разведочный анализ временных рядов и, в частности, проверять предположение о том, что распределение данных есть нормальное. Он четко показал, что ни о каком распределении Гаусса не может быть и речи, скорее имеет место гамма-распределение с параметрами, существенно отличающимися его от нормального.

Учет зависимости. Если тесты для анализа данных приняты, то можно обратиться к методам управления обработкой исходными данными и попытаться описать поведение используемых статистик для условно-реальных данных. При этом из-за наличия взаимосвязи элементов временного ряда непосредственное использование классических методов управления обработкой данных (бутстреп, складного ножа, перепроверка, тесты перестановки) подчас невозможно. Поэтому допустимо обратиться к образованию выборок путем повторного случайного пропуска отдельных элементов данных в предположении, что данная процедура несущественно искажает статистический вывод [8]. Но если принятие решений относительно экспериментальных результатов включает альтернативу, то применение любого из перечисленных методов управления обработкой, скорее всего, станет недопустимым: реальный временной ряд остается фактически единственным многомерным наблюдением.

Имея возможность моделировать анализируемые данные, можно все же добиться определенных результатов. Вернемся к анализу монотонного тренда с помощью критерия числа ступенек для выходного потока системы $M/M/2$. Выбор иллюстративных моделируемых систем опирается на соотношение (1) и для упрощения следует правилам: параметры p_1 и μ принимаются базовыми и $p_1 = 0,5$, $\mu = 1$; тогда λ выбирается как $C_\lambda \cdot 2\mu/(2 - p_1^2)$, где C_λ — индикатор стабильности работы системы и увеличение его значения приводит к росту «глубины» зависимости. Последнее следует из экспериментальных результатов, часть из которых приведена в табл. 1, в которой в 1-й строке приведены в процентах округленные значения:

$$T(j) = 1 - \sum_{i=1}^j \tau_N(i) \text{ при } N = 10^4,$$

в строках 2–6 — соответствующие эмпирические значения, найденные в ходе 10^3 экспериментов по моделированию системы $M/M/2$, а также серым выделены клетки в случае, когда при 1%-ном уровне значимости принимается

Таблица 1 Зависимость распределения числа ступенек от «глубины» зависимости

№	C_λ , %	Средняя «глубина»	$T(12)$	$T(13)$	$T(14)$	$T(15)$	$T(16)$
1	—	—	26	17	10	6	3
2	10	2	28	19	11	6	2
3	50	19	28	19	12	8	6
4	60	35	32	22	16	11	7
5	90	407	68	63	58	52	47
6	99	2021	75	75	74	73	72

гипотеза о равенстве эмпирической и теоретической частот при альтернативной гипотезе о том, что реальные ошибки критерия монотонного тренда на основе числа ступенек значимо превосходят теоретические значения, посчитанные для случайной выборки.

Представленные результаты иллюстрируют важные для анализа данных моменты:

- увеличение C_λ приводит к росту «глубины» зависимости наблюдаемого временного ряда (это сопряжено с удлинением очереди в системе);
- при относительно малых значениях «глубины» (до 19 согласно 2-й строке табл. 1) и умеренных значениях уровня значимости (например, 10% согласно столбцу $T(14)$) можно пользоваться базовым критерием числа ступенек без каких-либо коррективов; указанные значения 19 и 10 можно незначительно изменять в нужную сторону (соответственно увеличивать и уменьшать) без особенного риска;
- процесс прореживания можно проводить частично, не добиваясь нулевой «глубины» корреляции.

Завершим рассматриваемый пример иллюстрацией решения задачи выявления тренда. В качестве объекта анализа выступал набор $\{V_i\}$ времен пребывания в системе заданий с номерами $i = 1, \dots, 5 \cdot 10^4$, которая разбивалась на 5 фрагментов фиксированной длины из последовательных наблюдений. Для проверки нулевой гипотезы H_0 об отсутствии тренда применялся критерий числа ступенек для прореженных временных рядов для каждого фрагмента в отдельности при заданном уровне значимости, далее это $\alpha = 0,1$. Распределение числа ступенек дискретно и использовалось для значений объемов выборки, принимавших небольшие значения. Поэтому в каждом отдельном эксперименте находился критический уровень значимости критерия числа ступенек, который сравнивался с α , и принималось решение: подтвердить/опровергнуть H_0 . Подобный анализ (генерация всей последовательности $\{V_i\}$, принятие решения относительно наличия тренда для каждого фрагмента в отдельности) повторялся 10^3 раз. В результате получалась оценка α^* для α , относительно которой можно было уже формально делать вывод о ее равенстве α (значимый вывод об отсутствии тренда в данных

Таблица 2 Оценка (в %) критического уровня значимости критерия числа ступенек для первых фрагментов

№ фрагмента	Тип системы	
	$C_\lambda = 0,9$, система стабильна	$C_\lambda = 1,1$, система нестабильна
1	21	100
2	15	100
3	10	—
4-5	Повторение в принципе результатов для 3-го фрагмента	

фрагмента) или превышении над α (значимый вывод о наличие тренда в них). Соответствующие результаты представлены в табл. 2.

Содержание табл. 2 указывает на следующее:

- для $C_\lambda = 0,9$ и 1-го фрагмента явно предпочтительней выглядит вывод о наличии тренда, что соответствует поведению стабильной системы в начале своего функционирования; для остальных фрагментов обоснованно подтверждается гипотеза об отсутствии тренда;
- факт нестабильности для $C_\lambda = 1,1$ выявляется очевидным образом; но предложенный метод дает сбой для 3-го и последующих фрагментов из-за того, что после прореживания просто не остается данных для обработки.

Если метод прореживания становится неработоспособным, то можно отказаться от моделирования анализируемого временного ряда $\{V_{i^*}\}$ путем выбора с помощью прореживания отдельных элементов всей последовательности $\{V_i\}$ и перейти к последовательному получению элементов $\{V_{i^*}\}$ один за другим, заново генерируя $\{V_i\}$ полностью и отбирая по номеру i^* нужный элемент. Понятно, что рекомендовать такой способ как универсальный не приходится, но как средство верификации других решений его надо обязательно иметь в виду.

Обобщение модели. Поставив задачу выявления изменений характеристики исследуемого процесса, оправданно брать за основу модели данных с минимально возможным объемом априорных предположений. Но это не исключает попыток обратиться к более сложным, комплексным подходам. Ярким примером таковых служит класс параметрических моделей, описывающих нестационарные ряды с зависимостями в виде ARIMA-процесса (Autoregressive Integrated Moving Average). История зарождения данного подхода и его современное изложение даны в разд. 4.1.2 работы [9], примером широкого применения модели ARIMA при прогнозировании только по одному специфическому вопросу распространенности/заболеваемости COVID-19 может служить 21 ссылка на публикации в [10].

Обычная запись модели — $ARIMA(p, d, q)$, где p и q — порядки моделей соответственно авторегрессии и скользящего среднего для описания стационарного ряда, который должен получиться после d преобразований исходных данных.

Преобразования заключаются во взятии разностей первого порядка элементов ряда. Таким образом, ключевым элементом ARIMA-методики становится обнаружение нестационарности, для чего в рамках принятой общей линейной модели обычно применяется тест Дики–Фуллера (DF-тест) [11]. Это один из тестов на единичные корни, который опирается на представление

$$Y_t = \rho Y_{t-1} + e_t, \quad t = 1, 2, \dots; \quad Y_0 = 0 \quad (2)$$

и заключается в проверке нулевой гипотезы о равенстве единице коэффициента ρ . Поскольку случай «взрывных» процессов исключается, то тест становится односторонним (альтернативной служит гипотеза о том, что коэффициент ρ меньше единицы). Стьюдентизированная статистика DF-теста принимает привычную форму, но ее распределение иное (необходимые значения получены методом моделирования и представлены в приложении А работы [11]). Если значение статистики лежит левее критического значения при данном уровне значимости, то нулевая гипотеза о единичном корне отклоняется и процесс признается стационарным, в противном же случае — нестационарным (интегрированным). Важно обращать внимание на нюансы применения критерия. Например, более практичное предположение (оно используется при центрировании исходных данных) в виде

$$Y_t = (1 - \rho)\mu + \rho Y_{t-1} + e_t$$

приводит к иной таблице критических значений, чем в случае (2).

Если для DF-теста повторить эксперименты, подобные тем, что привели к результатам табл. 2, то получим крайне примечательные выводы:

- все решения о стационарности принимаются безвариантно (стопроцентная оценка вероятности принятия конкурирующей гипотезы), что при $C_\lambda = 0,9$ и $d = 0$ совершенно естественно, так как речь идет о стабильной системе, а в остальных случаях, когда $d = 1$, демонстрирует эффективность применения оператора разности первого порядка;
- однозначное отклонение гипотезы о стационарности при $C_\lambda = 1,1$ и $d = 0$ подтверждает эффективность DF-теста, но стопроцентные решения о стационарности при $C_\lambda = 0,9$ и $d = 0$ порождают сомнения в статистической корректности применения этого критерия (для сравнения генерировались ряды согласно модели авторегрессии первого порядка с корнем, меньшим 1, для которых в случае пяти последовательных фрагментов получились значения 11%, 10%, 9%, 11% и 10%, т. е. оценки заданного уровня значимости в 10%).

Таким образом, привлечение сложной ARIMA-модели, с одной стороны, обогащает круг решаемых задач (наряду с характеристикой наблюдаемого процесса типа «с трендом или без» появилась возможность его моделировать, например с прогностической целью), но с другой — порождает множество проблем. Главная из них — это проверка адекватности принятой модели данных, поверхностная реализация которой может привести к неполным, а возможно, и неправильным

выводам: например, если критерий числа ступенек выявил неординарное поведение выходной последовательности в начале функционирования системы $M/M/2$ (см. выводы из результатов табл. 2), то DF-тест на это «не обратил внимание». Далее, ARIMA-модель — параметрическая, и решение соответствующих задач оценивания провести строго обычно не удастся: например, процедура выбора значения d является последовательной и реально статистически не обоснована; оценивание отдельных значений p и q основывается на автокорреляциях, выборочные свойства которых до конца не известны; выбор p и q в совокупности фактически осуществляется перебором на основе некоторой «разумности», трудно поддающейся формализации. И, наконец, для исследования возможности использования общей линейной модели и определения конкретного типа этой модели должны анализироваться корни основанного на значениях автокорреляции характеристического полинома $\varphi(z)$, для поиска которых применяются итерационные методы (в частности, метод Ньютона). При этом могут возникнуть значительные трудности. Основная причина здесь состоит в том, что степень анализируемых полиномов может оказаться достаточно высокой. Кроме того, свойства корней полинома $\varphi(z)$ таковы, что найденное удовлетворительное приближение для корня z_0 может не подойти в качестве даже начального приближения для z_0^{-1} , хотя z_0^{-1} также должен быть корнем $\varphi(z)$.

Перечисленные проблемы становятся особенно острыми при необходимости решать задачу анализа временных рядов в автоматическом режиме.

3 Заключение

Популярной для описания зависимости от времени стала модель временного ряда, включающего тренд и набор взаимосвязанных элементов с нулевым средним. Хотя были предприняты значительные усилия для разработки тестов в простом с точки зрения практики случае линейного тренда при наличии последовательной корреляции, соответствующие процедуры анализа данных оказались менее чем удовлетворительны для всего множества параметров, характеризующих зависимость элементов временного ряда.

Во многих прикладных областях оценка характера тенденций и статистической значимости полученных решений остается проблемой. Во-первых, редко когда удастся априори определиться с видом тренда (например: линейный, экспоненциальный и т. д.), тем более в долгосрочном смысле. Во-вторых, данные чаще всего в силу своей природы распределены негауссово. В-третьих, если изменчивость практически всегда может быть идентифицирована, то ключевой вопрос о том, можно ли считать ее статистически значимой, остается полностью не закрытым. В отличие от обычно выдвинутых предположений в условиях случайной выборки, временной ряд может дополнительно иметь последовательную зависимость — либо краткосрочную, либо долгосрочную, что существенно влияет на статистическую значимость соответствующего вывода относительно тренда.

В-четвертых, известные модели временных рядов с трендами в большей степени отвечают разведочному анализу данных, порождая методы типа «удостовериться, является ли временной ряд стационарным», «подобрать порядок разностного оператора, который приводит к стационарному ряду», «построить графики автокорреляционных функций и с их помощью определить значения параметров модели», «выполнить подгонку параметров для идентификации модели». Подобные указания не формализованы полностью, подразумевают дальнейшее развитие и конкретизацию соответствующих статистических процедур.

Рассмотренная модель монотонного тренда решает проблему общности модели тренда и недостаточности предположения о гауссовском распределении данных. Предложенные для применения приемы декорреляции частично обеспечивают оценивание значимости статистического вывода о наличии тренда, но ценой потери данных. Правда, в случае возможности повторного воспроизведения данных это препятствие преодолевается. Заманчиво выглядит ARIMA-модель, но подчас соответствующие работы воспринимаются только как благие намерения (см., например, [1]). В любом случае ясно, что на данный момент этот подход должен развиваться с привязкой к постановке конкретных задач практики и реальных условий их решения.

Литература

1. *Rafsanjani M. K., Rezaei A., Shahraki A., Saeid A. B.* QARIMA: A new approach to prediction in queue theory // *Appl. Math. Comput.*, 2014. Vol. 244. P. 514–525. doi: 10.1016/j.amc.2014.06.108.
2. *Кривенко М. П.* Распределение статистики отношения правдоподобия для выявления монотонного тренда // *Системы и средства информатики*, 2021. Т. 31. № 4. С. 27–37. doi: 10.14357/08696527210403.
3. *Miles R. E.* The complete amalgamation into blocks, by weighted means, of a finite set of real numbers // *Biometrika*, 1959. Vol. 46. Iss. 3-4. P. 317–327. doi: 10.2307/2333529.
4. *Roy A., Falk B., Fuller W. A.* Testing for trend in the presence of autoregressive error // *J. Am. Stat. Assoc.*, 2004. Vol. 99. No. 468. P. 1082–1091. doi: 10.1198/016214504000000520.
5. *Mair A., Fares A.* Time series analysis of daily rainfall and streamflow in a volcanic dike-intruded aquifer system, O’ahu, Hawai’i, USA // *Hydrogeol. J.*, 2011. Vol. 19. Iss. 4. P. 929–944. doi: 10.1007/s10040-011-0740-3.
6. *Brill P. H., Green L.* Queues in which customers receive simultaneous service from a random number of servers: A system point approach // *Manage. Sci.*, 1984. Vol. 30. No. 1. P. 51–68. doi: 10.1287/mnsc.30.1.51.
7. *Doornik J., Hansen H.* An omnibus test for univariate and multivariate normality // *Oxford B. Econ. Stat.*, 2008. Vol. 70. Iss. S1. P. 927–939. doi: 10.1111/j.1468-0084.2008.00537.x.
8. *Кривенко М. П.* Критерии выбора размерности модели факторизации // *Информатика и её применения*, 2023. Т. 17. Вып. 2. С. 50–56. doi: 10.14357/19922264230207.

9. *Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M.* Time series analysis: Forecasting and control. — 5th ed. — Hoboken, NJ, USA: Wiley, 2015. 712 p.
10. *Ceylan Z.* Estimation of COVID-19 prevalence in Italy, Spain and France // *Sci. Total Environ.*, 2020. Vol. 729. P. 1–7. doi: 10.1016/j.scitotenv.2020.138817.
11. *Fuller W. A.* Introduction to statistical time series. — 2nd ed. — Hoboken, NJ, USA: Wiley, 1996. 698 p.

Поступила в редакцию 05.06.23

TIME SERIES MONOTONIC TREND ANALYSIS

M. P. Krivenko

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The problem of identifying changes in the characteristics of the time series under study during the observation period is considered. Almost always, the trend arises against the background of the statistical dependence of the elements of the time series. The dependence of individual observations becomes a nuisance factor. When constructing assumptions about changes, preference is given to models of a general nature: the trend of probabilistic characteristics is a monotonic function of an unknown form from the observation number (monotonic trend). The need to take into account the combination of both operating factors in the time series model and the need to obtain workable methods of data processing lead to the following scheme of actions: to take as a basis the already developed procedures, then, if possible, to adjust the conditions for their correct application to the required ones, and, finally, to use adapted options. The analysis of methods for processing the nuisance factor is carried out: neutralization of dependence, accounting for dependence, and generalization of the time series model. As an example, the problem of monitoring the stable operation of a two-processor task processing system with a random selection of the number of required processors is considered. The possibilities and limitations of the proposed methods are demonstrated.

Keywords: monotone trend; decorrelation; ARIMA models

DOI: 10.14357/08696527230302

EDN: WVGUVX

References

1. Rafsanjani, M. K., A. Rezaei, A. Shahraki, and A. B. Saeid. 2014. QARIMA: A new approach to prediction in queue theory. *Appl. Math. Comput.* 244:514–525. doi: 10.1016/j.amc.2014.06.108.
2. Krivenko, M. P. 2021. Raspredelenie statistiki otnosheniya pravdopodobiya dlya vyavleniya monotonogo trenda [Distributions of likelihood ratio statistics for monotone trend detection]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(4):27–37. doi: 10.14357/08696527210403.

3. Miles, R. E. 1959. The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika* 46(3-4):317–327. doi: 10.2307/2333529.
4. Roy, A., B. Falk, and W. A. Fuller. 2004. Testing for trend in the presence of autoregressive error. *J. Am. Stat. Assoc.* 99(468):1082–1091. doi: 10.1198/016214504000000520.
5. Mair, A., and A. Fares. 2011. Time series analysis of daily rainfall and streamflow in a volcanic dike-intruded aquifer system, O’ahu, Hawai’i, USA. *Hydrogeol. J.* 19(4):929–944. doi: 10.1007/s10040-011-0740-3.
6. Brill, P. H., and L. Green. 1984. Queues in which customers receive simultaneous service from a random number of servers: A system point approach. *Manage. Sci.* 30(1):51–68. doi: 10.1287/mnsc.30.1.51.
7. Doornik, J., and H. Hansen. 2008. An omnibus test for univariate and multivariate normality. *Oxford B. Econ. Stat.* 70(S1):927–939. doi: 10.1111/j.1468-0084.2008.00537.x.
8. Krivenko, M. P. 2023. Kriterii vybora razmernosti modeli faktorizatsii [Criteria for choosing the factorization model dimensionality]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):50–56. doi: 10.14357/19922264230207.
9. Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time series analysis: Forecasting and control*. 5th ed. Hoboken, NJ: Wiley. 712 p.
10. Ceylan, Z. 2020. Estimation of COVID-19 prevalence in Italy, Spain and France. *Sci. Total Environ.* 729:1–7. doi: 10.1016/j.scitotenv.2020.138817.
11. Fuller, W. A. 1996. *Introduction to statistical time series*. 2nd ed. Hoboken, NJ: Wiley. 698 p.

Received June 5, 2023

Contributor

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

ДИСПЕТЧЕРИЗАЦИЯ В ЧАСТИЧНО НАБЛЮДАЕМЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ КОНЕЧНОЙ ЕМКОСТИ С ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ*

М. Г. Коновалов¹, Р. В. Разумчик²

Аннотация: Рассматривается новая задача оптимального централизованного управления выбором ресурсов в простейших моделях систем добровольных вычислений конечной емкости. Система моделируется на сильно агрегированном уровне, где она, по сути, вырождается в частично наблюдаемую стохастическую систему с параллельным обслуживанием с одним диспетчером, размещающим задания по однопроцессорным серверам без какой-либо информации об их текущем состоянии, причем очереди на серверах могут иметь конечную емкость. Частичная наблюдаемость подразумевает, что диспетчер осуществляет управление в отсутствие обратной связи с системой. Требуется предъявить стратегию, наилучшим образом оптимизирующую либо предельные значения среднего времени отклика и вероятности потери, либо заданную функцию от них. Описывается метод порождения соответствующих стратегий, реализующих известную идею диспетчеризации по предыстории в системах распределенных вычислений.

Ключевые слова: системы с параллельным обслуживанием; диспетчеризация; стратегии размещения заданий; управление при неполном наблюдении; программное управление

DOI: 10.14357/08696527230303

EDN: XUVVNH

1 Введение

В настоящее время во всем мире продолжают создавать и широко использовать инфраструктуры распределенных вычислений различного масштаба и назначения. Среди обилия проблем, возникающих в связи с задачами синтеза и эксплуатации, имеется по крайней мере одна универсальная, присущая всем оттенкам концептуальных подходов и примерам практических воплощений подобных систем. Эта проблема заключается в поиске ответа на вопрос: как следует распоряжаться вычислительными ресурсами системы? Или, другими

*Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mkonovalov@ipiran.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, rrazumchik@ipiran.ru

словами: куда, в какие моменты, в какой очередности и сколько следует направлять заданий, чтобы максимально удовлетворить потребителей и/или владельцев ресурсов? Указанная проблема многопланова и требует для своего решения привлечения различных подходов с учетом особенностей рассматриваемых систем (принципов их функционирования, применяемых технологий и т. п.).

В данной статье речь пойдет о задаче оптимального управления выбором ресурсов (т. е. о задаче определения наилучшей с точки зрения заданного критерия стратегии выбора ресурсов для поступающих заданий) в частично наблюдаемых стохастических системах конечной емкости с параллельным обслуживанием¹. Рассматривается простейший с точки зрения структуры случай, когда конечная совокупность параллельно и независимо друг от друга работающих обслуживающих ресурсов (далее — серверов) обязательно одинаковой производительности выполняет задания, направляемые на них единственным диспетчером, не имеющим очереди для хранения заданий. Частичная наблюдаемость подразумевает, что диспетчер, осуществляя выбор сервера для очередного задания, может руководствоваться только:

- априорной информацией о системе, т. е. распределениями времен между поступлениями заданий, распределениями объемов заданий, статическими характеристиками серверов (производительность, емкость очереди, дисциплина обслуживания очередей и т. п.);
- значениями моментов времени поступления предыдущих заданий и информацией об уже принятых решениях.

Описанная схема — пожалуй, простейшая из содержательных² схем для моделирования специального класса реально функционирующих систем распределенных вычислений, так называемых систем добровольных вычислений, имеющих централизованную архитектуру (см. таксономию в [3]). Сформулированная задача связана с процессом оперативного управления ими, требующим многократного и быстрого принятия решений. В научной литературе она встреча-

¹В связи с выдвинутым тезисом о многоплановости отметим, что обозначенная задача принципиально отличается от таких важных для рассматриваемых систем вопросов, как оптимальное «разбиение» заданий и выбор оптимальных дисциплин обслуживания очередей «внутри» ресурсов. Кроме того, ее решение, вообще говоря, зависит от ответов на эти вопросы. Задача же выбора оптимального алгоритма управления всей системой настолько сложна (см., например, [1]), что для нее в достаточно общей форме в настоящее время вообще нет приемлемых конструктивных методов решения. Размерность и сложность структуры получающейся математической задачи делает целостное решение ее практически невозможным. Естественным выходом из такого положения (которое также характерно для проблематики сетей связи) видится последовательная оптимизация упомянутых частных задач со всеми вытекающими новыми сложностями (как, например, неясные условия сходимости процесса последовательных приближений; см. более общий, системный взгляд в [2]).

²Ее использование предоставляет быстрый и дешевый способ получения оценок производительности на стадии проектирования, что способствует принятию обоснованных решений относительно аппаратного и программного исполнения систем.

ется и под другими названиями: алгоритм распределения ресурсов/ размещения заданий, стратегия маршрутизации, диспетчеризация. Далее эти словосочетания используются как синонимы.

Добровольные вычисления — известный подход к решению вычислительно сложных задач. Согласно [3], волна интереса к ним возникла в конце 1980-х гг. и, судя по публикациям в научной литературе, не ослабла до сих пор: продолжают создаваться новые проекты, нацеленные на решение широкого круга важных практических задач¹. По данным [7], к BOINC — наиболее известной платформе для организации добровольных вычислений — сейчас подключено более 150 тыс. серверов. Широкое применение систем добровольных вычислений выдвигает на первый план вопрос повышения их производительности. Подобно тому как в сетях связи решение аналогичного вопроса связывают с оптимизацией маршрутных алгоритмов (и опыт эксплуатации реальных сетей это подтверждает; см., например, обзор в [8, разд. 4]), здесь ключевая роль отводится алгоритму распределения заданий. Поэтому на сегодняшний день для систем добровольных вычислений уже разработано большое число диспетчеризаций, позволяющих оптимизировать классические целевые функционалы (например, стационарное среднее время отклика/ожидания, среднее число обработанных за заданное время заданий). Эффективность этих диспетчеризаций обоснована в разной степени² и в разной мере учитывает техническую сторону вопроса.

Достаточно полный обзор диспетчеризаций, известных к 2019 г. и доведенных до практической реализации, приведен в [3, 9, 10] (см. также [11]). В [3, subsection 2.3] предложена классификация, согласно которой диспетчеризация в системе добровольных вычислений является либо информационно-ориентированной, т. е. учитывающей доступную информацию о предыстории работы и/или текущем состоянии ресурсов, либо «наивной», распределяющей задания вслепую, без какой-либо динамической информации о ресурсах. Главными представителями последнего класса выступают статические стратегии³. Несмотря на то что от их использования нельзя ожидать большого выигрыша, именно они стали стандартом де-факто: хотя бы одна такая стратегия реализуется в планировщике каждой системы распределенных вычислений. Причины этого — их универсальность, относительная простота реализации и легкое масштабирование. Действительно, самая известная из статических стратегий — это случайный выбор⁴ (далее — RND). Если минимизируется стационарное среднее

¹ Например, DreamLab, SiDock@home, Gaia@home, BOID; см. также [4–6].

² Например, известны решения, использующие приемы так называемой эволюционной оптимизации для нахождения эффективного правила распределения заданий (см., например, Table 1 в [3]).

³ Стратегия считается статической, если номер сервера для любого задания может быть определен до начала управления.

⁴ В зарубежной литературе — PAP (Probabilistic Allocation Policy), или RND (Random), или BS (Bernoulli Splitting). Подробное описание упоминаемых далее диспетчеризаций для систем распределенных вычислений можно найти, например, в [12].

время отклика в стандартных условиях, то, выбирая значения параметров стратегии RND таким образом, чтобы нагрузка балансировалась между серверами пропорционально их производительности, можно получить простой алгоритм приемлемого качества (которое иногда можно оценить и аналитически), не требующий дополнительных вычислений. Большого выигрыша можно добиться, если использовать те же значения¹ параметров в других статических стратегиях — так называемых программных² (далее — PROG), имеющих небольшую вычислительную сложность (см., в частности, алгоритм SG в [13, с. 184] и [14]). Попытки понять, где находится оптимум при управлении выбором ресурсов в отсутствие наблюдений за состояниями ресурсов, привели к разработке принципиально нового типа диспетчеризации — диспетчеризации по предыстории (далее — AA, от *англ.* Arrival Aware), превосходящей, судя по модельным экспериментам, как RND, так и PROG. Хотя известные ее реализации (см., например, [11, 14–16]) по трудоемкости обычно уступают статическим алгоритмам, получающийся в стандартных условиях выигрыш полностью оправдывает ее применение³.

Помимо практической выгоды интерес к задачам оперативного управления и, в частности, к задаче оптимального управления выбором ресурсов объясняется их хорошей формализуемостью. Это обстоятельство позволяет сводить такие задачи к некоторым экстремальным, которые, хотя и редко поддаются точному анализу, обычно могут быть решены приближенно. Однако современные исследования практических проблем построения систем распределенных вычислений [3, 9, 10] порождают и принципиально новые постановки, для которых существующий арсенал прикладных математических методов оказывается недостаточным.

Эта статья посвящена одной из таких постановок, которая возникает при попытках аналитически подступиться к вопросу⁴ оптимального централизованного управления выбором ресурсов в системах добровольных вычислений конечной емкости. Система рассматривается на сильно агрегированном уров-

¹Если же значения параметров стратегии RND выбраны оптимальным образом, то их использование в подходящих программных стратегиях может привести к лучшему значению целевого функционала. Более подробный сравнительный анализ см. в [8, с. 29–36]. Напомним, что циклический выбор является частным случаем программной стратегии и оптимален в системах из однородных серверов.

²Эти стратегии параметризуются бесконечными последовательностями $\{a_1, a_2, \dots, a_n, \dots\}$, где a_i означает, что i -е по счету задание направляется на сервер с номером a_i .

³Отметим, что до сих пор неизвестно, насколько далеко можно продвинуться в задаче диспетчеризации, оставаясь в классе «наивных» стратегий: в общем случае оптимальные значения классических целевых функционалов остаются неизвестными.

⁴Решение этого вопроса в первую очередь необходимо для задач системного анализа: получения оценок и рекомендаций на стадии проектирования, выработки наилучших способов эксплуатации системы. С прикладной точки зрения вопрос этот малоинтересен: информационно-ориентированные диспетчеризации позволяют при распределении заданий учитывать ограничения на объем памяти ресурсов «волонтеров».

не, где она, по сути, вырождается в частично наблюдаемую стохастическую систему с параллельным обслуживанием с одним диспетчером, размещающим задания по однопроцессорным серверам без какой-либо информации об их текущем состоянии, причем очереди на серверах могут иметь конечную емкость. Следствие последнего условия — потери заданий, высокий показатель которых обыкновенно трактуется как низкое качество функционирования системы.

Таким образом, для нахождения наилучшей диспетчеризации здесь необходимо решать задачу либо многокритериальной оптимизации (например, минимизировать одну временную и одну вероятностную характеристику поступающих в систему заданий), либо однокритериальной, но соединяющей в себе несколько целевых функций. В обоих случаях нельзя обойтись без вычисления значений целевой функции, связанной с потерями заданий, что в рассматриваемых условиях частичной наблюдаемости¹ сделать невозможно. Это обстоятельство затрудняет применение известных статических алгоритмов: их параметры и требуемые для реализации величины не могут быть рассчитаны.

В этой статье предложен конструктивный метод для порождения диспетчеризаций, реализующих для рассматриваемых систем предложенную в [15] идею управления потоками заданий по предыстории принятых решений.

Как и в [14], в основе метода лежит использование для порождения решений виртуальных вспомогательных процессов, зависящих от неизвестных параметров и синхронизованных по моментам поступления заданий с основной системой. Ввиду того что имеющаяся по предположению априорная информация дает возможность осуществлять имитацию траектории системы, значения неизвестных параметров могут быть подобраны. В силу сложности характеристики таких алгоритмов не удастся оценить аналитически. Модельные эксперименты показывают, что они успешно конкурируют со статическими алгоритмами (RND и PROG), применение которых требует предварительной оценки параметров² на той же имитационной модели. Это свойство вместе с тем обстоятельством, что новые алгоритмы не требуют специальных ограничений ни на входящий поток, ни на процесс обслуживания, свободны от вычислительных недостатков и обходятся для своей настройки оценкой существенно меньшего числа параметров, дают основания назвать их лучшими для рассматриваемых систем в условиях частичного наблюдения.

В следующем разделе приводится формальная постановка задачи. Затем в разд. 3 излагается метод создания диспетчеризаций для рассматриваемых систем. Несколько примеров, иллюстрирующих характеристики новых алгоритмов, приведены в разд. 4. В разд. 5 резюмируются основные результаты и затрагиваются некоторые открытые вопросы.

¹Эквивалентных (в терминах теории оптимального управления динамическими стохастическими системами) отсутствию информации, отсутствию обратной связи.

²Причем число неизвестных параметров совпадает с размерностью системы.

2 Диспетчеризация в системах добровольных вычислений как задача распределения потока случайных объемов вычислений в частично наблюдаемых стохастических системах конечной емкости с параллельным обслуживанием

Рассматривается система, состоящая из N работающих параллельно однопроцессорных серверов, пронумерованных числами от 1 до N без повторений. Сервер с номером i характеризуется производительностью $R_i > 0$ и доступным объемом памяти $C_i > 0$ для хранения заданий. Выбор очередного задания на обслуживание на каждом сервере происходит в порядке поступления. В систему поступает единственный поток однородных заданий, причем интервалы $\theta_1 = t_2 - t_1, \theta_2 = t_3 - t_2, \dots$ между моментами t_1, t_2, \dots поступления заданий имеют одинаковую функцию распределения F . Номер сервера, на который направлено n -е по счету задание, далее обозначается через $y_n \in \{1, \dots, N\}$. Значение y_n определяется диспетчеризацией. Предполагается, что функция распределения B объема задания n имеет конечное среднее значение b . Объем задания n далее обозначается через v_n . Будучи направленным на сервер, он принимается к выполнению либо полностью, если на сервере достаточно свободного места, либо частично. В последнем случае отклоненная часть задания (далее — r_n) считается потерянной. Обозначая через a_n принятую на обслуживание часть задания, а через $x_{y_n}(n)$ — объем незаконченной работы на сервере y_n в момент поступления n -го по счету задания в систему, имеем $a_n = \min\{C_{y_n} - x_{y_n}(n), v_n\}$, $r_n = v_n - a_n$.

Предполагается, что каждое поступившее задание должно быть немедленно направлено диспетчером на один из серверов. Цель диспетчеризации — минимизировать стационарное среднее время пребывания задания в системе и стационарную вероятность потери задания или увязывающий оба показателя заданный функционал. При принятии решения относительно задания n помимо функций распределения F и B , значений $R_1, \dots, R_N, C_1, \dots, C_N, t_n$, сведений о дисциплинах обслуживания на серверах диспетчеру известна только предыстория принятых им решений y_1, \dots, y_{n-1} и моменты времени t_1, \dots, t_{n-1} , в которые эти решения принимались.

Сформулированная выше цель диспетчеризации остается неполной без указания того, каким образом рассчитываются два ключевых показателя производительности. Действительно, время пребывания обычно относят к заданию, т. е. рассматривают величину¹

$$T = \lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{V_n}{N},$$

где $V_n = (x_{y_n}(n) + a_n)/R_{y_n}$ — время пребывания в системе той части задания n , которая принята на обслуживание. Поскольку небольшое время пребывания

¹Здесь и далее предполагается, что пределы существуют и детерминированы почти наверное относительно мер, порожденных рассматриваемыми диспетчеризациями.

может достигаться за счет того, что бóльшая часть задания потеряна, то имеет смысл рассматривать также величину

$$T^* = \lim_{N \rightarrow \infty} \frac{\sum_{n=1}^N V_n}{\sum_{n=1}^N a_n},$$

которую можно интерпретировать как предельное среднее время пребывания единицы объема, принятого на обслуживание. Уровень потерь также может быть оценен с помощью двух в равной степени пригодных показателей:

$$Q = \lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{r_n}{N v_n}; \quad Q^* = \lim_{N \rightarrow \infty} \frac{\sum_{n=1}^N r_n}{\sum_{n=1}^N v_n}.$$

Величина Q отражает предельную среднюю долю потерь, приходящуюся на одно задание, тогда как значение Q^* указывает на предельную среднюю долю потерь от всего объема заданий, поступивших в систему.

Показатели времени пребывания и уровня потерь, безусловно, взаимосвязаны. Однако в рассматриваемой постановке характер этой зависимости интуитивно не очевиден. Например, можно предположить, что распределение заданий, при котором уменьшается время пребывания и, следовательно, освобождается больше свободного места на серверах, уменьшает также количество потерь. С другой стороны, уменьшить время пребывания можно, увеличив объем потерь. Поэтому естественно оценивать качество работы системы сверткой этих двух показателей, придав ей традиционный стоимостный смысл. Определим одношаговые затраты l_n , связанные с одним заданием, полагая

$$l_n = \alpha V_n + \beta r_n. \tag{1}$$

Здесь константа α означает цену за единицу времени пребывания, а константа β соответственно — цену за единицу потерь. Тогда предельные средние затраты L можно определить как

$$L = \lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{l_n}{N}.$$

3 Метод порождения диспетчеризаций

Пусть наряду с процессом поступления и обслуживания заданий в рассматриваемой системе имеется еще $K \geq 1$ вспомогательных процессов, моделирующих точно такую же систему, т.е. с такими же серверами и пр.¹

¹Подчеркнем, что основной процесс отображает «реальный» процесс (хотя и может быть заменен имитационной моделью), в то время как вспомогательные процессы — чисто виртуальные компьютерные модели.

Начальные состояния всех процессов одинаковые. Функционирование вспомогательных процессов в точности копирует работу основного процесса в части моментов поступления, размещения заданий, дисциплины обслуживания, но различается объемами заданий. Кроме того, различие между основным и вспомогательными процессами заключается в характере наблюдений. Если для основного процесса наблюдения по условию ограничены моментами поступления заданий и совершенными управлениями, то вспомогательные процессы полностью наблюдаемы.

Зададимся положительными числами $D^{(1)}, \dots, D^{(K)}$ и будем считать, что все задания, поступающие на серверы k -го вспомогательного процесса (идентичные с серверами основного процесса), имеют одинаковый объем $D^{(k)}$, $1 \leq k \leq K$. Обозначим через $z_i^{(k)}(n)$ время, необходимое для выполнения всей незаконченной работы, имеющейся на сервере i вспомогательного процесса с номером k в момент поступления n -го задания, но без его учета. Тогда, вне зависимости от дисциплины обслуживания в основном процессе, значения $z_i^{(k)}(n)$ и $z_i^{(k)}(n-1)$ будут связаны рекурсией Линдли:

$$z_i^{(k)}(n) = \max \left\{ 0, z_i^{(k)}(n-1) + \delta_{i, y_{n-1}} D^{(k)} - \theta_{n-1} R_i \right\}, \quad n = 1, 2, \dots \quad (2)$$

Зафиксируем для каждого вспомогательного процесса функцию¹, позволяющую количественно оценивать в моменты поступления заданий состояние каждого его сервера с учетом всех имеющихся ограничений². Предпишем k -му диспетчеру направлять задания на его сервер, имеющий минимальную оценку³. Обозначим через $y_1^{(k)}, y_2^{(k)}, \dots$ решения в k -м вспомогательном процессе, порожденные принятой диспетчеризацией. Пусть теперь основной и K вспомогательных процессов проработали до некоторого момента t_n , $n = 1, 2, \dots$, поступления очередного задания в основном процессе. Этот момент принудительно считается моментом поступления и для всех вспомогательных процессов, однако объем нового задания определяется для каждого из них индивидуально. Для каждого вспомогательного процесса, исходя из известной для него оценочной функции, определяется номер сервера, на который отправляется его собственное новое задание. Решением y_n диспетчера основного процесса считается наиболее часто встречающееся число в наборе $\{y_n^{(1)}, \dots, y_n^{(K)}\}$. Далее работа основного и вспомогательных процессов протекает независимо друг от друга вплоть до следующего момента поступления задания в основном процессе.

¹ В силу того что вспомогательные процессы полностью наблюдаемы, в качестве оценочных функций могут выступать известные алгоритмы для систем с параллельным обслуживанием: JSQ, HJSQ(d), LWL, Myopic, SITA-E, SITA-V и др. (см. обзоры [12, 17]).

² Ограничения могут появиться, например, при сведении задачи многокритериальной оптимизации к однокритериальной многомерной задаче условной оптимизации.

³ Здесь и всюду далее возможная неоднозначность разрешается случайным выбором.

4 Некоторые результаты моделирования

Сложность решаемой задачи такова, что характеристики любой осмысленной диспетчеризации вне зависимости от предположений о распределениях F и B не представляется возможным оценить аналитически: при выборе сервера не наблюдаются даже показатели, подлежащие оптимизации. Поэтому в настоящее время для сравнения диспетчеризаций в рассматриваемых условиях альтернативы эксперименту, т. е. имитационному моделированию, не существует.

Остановимся на нескольких примерах, дающих представление о соотношениях между алгоритмами распределения заданий при минимизации предельных средних затрат L . При этом, однако, будем отслеживать также и показатели T , T^* , Q и Q^* .

В качестве эталонного алгоритма выберем алгоритм RND (см. разд. 1), который описывается набором (p_1, \dots, p_N) из N чисел — вероятностей p_n выбора для очередного задания сервера n .¹ Несмотря на то что проблема нахождения оптимального набора (p_1, \dots, p_N) хорошо известна, в условиях рассматриваемой задачи он не может быть вычислен² ни в многокритериальном случае, ни при условной оптимизации. Поэтому в экспериментах значения параметров оптимизировались на имитируемых траекториях.

Алгоритм на основе предложенного в предыдущем разделе метода реализует идею диспетчеризации по предыстории. Поэтому далее, как и в разд. 1, будем использовать для него аббревиатуру АА. Для создания алгоритма АА необходимо зафиксировать число вспомогательных процессов K , значения $D^{(1)}, \dots, D^{(K)}$ и оценочную функцию. Положим $K = 1$ и в качестве оценки — одношаговые затраты (1). Тогда задание с номером n , поступившее во вспомогательном процессе, направляется на сервер с номером $y_n^{(1)}$, который вычисляется по формуле³:

$$y_n^{(1)} = \operatorname{argmin}_{1 \leq i \leq N} \left\{ \alpha \frac{z_i^{(1)}(n) + \min \{C_i - z_i^{(1)}(n), D^{(1)}\}}{R_i} + \beta \left(D^{(1)} - \min \{C_i - z_i^{(1)}(n), D^{(1)}\} \right) \right\}, \quad (3)$$

¹Здесь выбор между рандомизированной и программной стратегиями в пользу первой объясняется тем, что для программной стратегии в общем случае не ясно, как находить оптимальную последовательность действий (впрочем, см. [18]). Обычно в качестве значений ее параметров берут оптимальные значения (p_1, \dots, p_N) для стратегии RND, и программная стратегия сводится просто к «максимальному расщеплению» потока по ресурсам и никак не учитывает ограничения на объем памяти ресурсов. Ограничения эти учтены в значениях (p_1, \dots, p_N) , т. е. в стратегии RND.

²Быть может, за исключением полностью экспоненциальных систем.

³Возможная неоднозначность всегда разрешается случайным выбором.

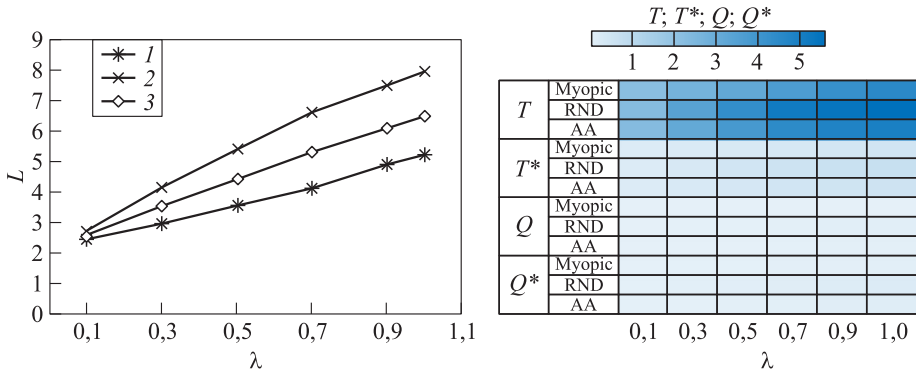


Рис. 1 Зависимость предельных средних затрат от интенсивности входящего потока: 1 — Муоріс; 2 — RND; 3 — АА

в которой значения $z_1^{(1)}(n), \dots, z_N^{(1)}(n)$ предварительно рассчитываются по (2). Поскольку $K = 1$, то решение y_n диспетчера основного процесса совпадает с $y_n^{(1)}$. Параметр $D^{(1)}$ — единственный неизвестный параметр диспетчеризации (3) при любом числе серверов, и подобрать его можно только на имитируемых траекториях. Для этого применимы адаптивные алгоритмы управления частично наблюдаемыми марковскими цепями (см. также [19, 20]) и алгоритмы эволюционной оптимизации¹.

С целью контроля вычислений добавим к рассмотрению информационно-ориентированную стратегию с решающим правилом для n -го задания

$$\operatorname{argmin}_{1 \leq i \leq N} \left\{ \alpha \frac{x_i(n) + \min \{C_i - x_i(n), v_n\}}{R_i} + \beta (v_n - \min \{C_i - x_i(n), v_n\}) \right\},$$

которую естественно назвать миопической (далее — Муоріс).

Будем рассматривать систему с заданиями экспоненциального объема (со средним $b = 10$), поступающими в виде пуассоновского потока с параметром λ к диспетчеру, распределяющему их по $N = 8$ серверам со следующими характеристиками производительности и объема памяти: $(R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8) = (1, 1, 1, 1, 2, 2, 2, 5)$; $(C_1, \dots, C_N) = (5, 10, 15, 20, 25, 30, 35, 40)$. Зависимость предельных средних затрат L от интенсивности входящего потока λ в такой системе иллюстрирует рис. 1. Предполагается, что $\alpha = \beta = 1$, т. е. предпочтения при минимизации L между временем пребывания и потерями нет. Значения параметров стратегий RND и АА при каждом значении λ (табл. 1) находились путем оптимизации на имитируемых траекториях.

¹ Впрочем, из-за особенностей задачи эти алгоритмы не сильно отличаются от ручного перебора.

Таблица 1 Значения параметров для стратегий RND и AA из рис. 1

Стратегия	Параметр	Интенсивность входящего потока λ					
		0,1	0,3	0,5	0,7	0,9	1,0
RND	p_1	0,0002	0,0001	0,0003	0,0001	0,1073	0,1290
	p_2	0,0002	0,0001	0,0001	0,0001	0,0369	0,0435
	p_3	0,0002	0,0001	0,0003	0,0002	0,0237	0,0268
	p_4	0,0002	0,0000	0,0002	0,0001	0,0190	0,0212
	p_5	0,0004	0,0407	0,0973	0,1245	0,1234	0,1212
	p_6	0,0004	0,0498	0,0984	0,1197	0,1118	0,1087
	p_7	0,0004	0,0555	0,0981	0,1156	0,1042	0,1010
	p_8	0,9980	0,8539	0,7054	0,6397	0,4738	0,4487
AA	$D^{(1)}$	17	16	15,5	16,5	17	17

Как видно из рис. 1, вне зависимости от того, насколько сильно загружена система (загрузка равна $2\lambda/3$, диспетчеризация на основе вспомогательных процессов превосходит рандомизированную, причем не только по критерию минимума свертки двух критериев, но и по каждому из них в отдельности. Вычислительные эксперименты показывают, что так обстоит дело и при самых общих предположениях об исходных параметрах системы: для большинства встречающихся на практике распределений входящего потока и объемов заданий алгоритмы, разработанные на основе метода разд. 3, равномерно лучше рандомизированной стратегии, даже при оптимальных значениях параметров последней. Выигрыш при разной загрузке неодинаков: в сравнении с RND обычно чем выше загрузка, тем больше выигрыш¹. Это наблюдение еще раз подчеркивает плодотворность предложенного метода диспетчеризации. Действительно, при высоких загрузках очереди на серверах не пусты большую часть времени. Наблюдаемый же выигрыш в сравнении со стратегией RND свидетельствует о том, что построенный по методу разд. 3 алгоритм достаточно чувствителен, чтобы уловить разброс значений незаконченной работы на серверах, возникающий за время между двумя последовательными поступлениями.

Все сказанное выше справедливо при условии, что значение неизвестного параметра $D^{(1)}$ диспетчеризации AA подобрано правильно. Это обстоятельство проиллюстрировано на рис. 2, из которого видно, что без оптимизации обойтись нельзя. Однако, как показывают модельные эксперименты, в каждой конкретной системе существует лишь единственное оптимальное значение $D^{(1)}$, причем в достаточно большой его окрестности классические целевые функционалы «выходят» на свои близкие к оптимальным значения². Поэтому, в отличие от

¹Однако в сравнении с хорошо настроенной стратегией PROG с ростом загрузки выигрыш уменьшается.

²Так обстоит дело и с алгоритмами, построенными по методу разд. 3, с $K > 1$ параметрами $D^{(1)}, \dots, D^{(K)}$.

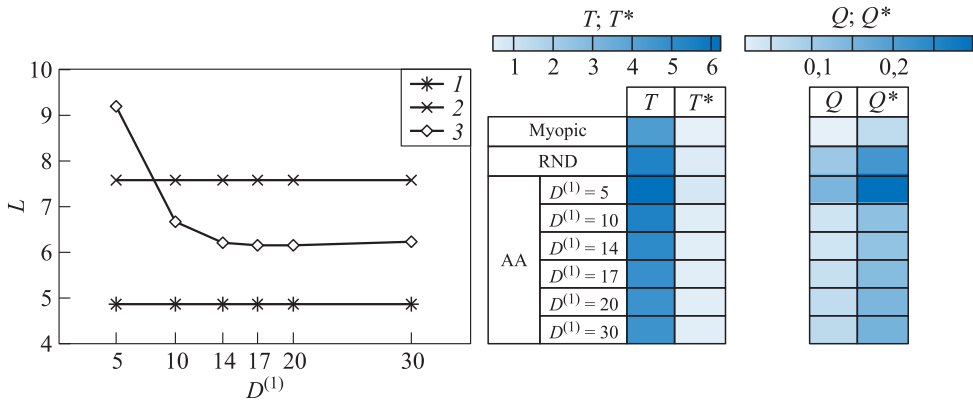


Рис. 2 Поведение диспетчеризации AA при $\lambda = 0,9, \alpha = \beta = 1$ и различных значениях параметра $D^{(1)}$: 1 — Муорис; 2 — RND; 3 — AA. Параметры стратегии RND взяты из соответствующего столбца в табл. 1

диспетчеризации RND, в которой определению подлежат $N - 1$ параметров, хорошая диспетчеризация AA может быть идентифицирована относительно просто. Важно учитывать, что, как известно, в общем случае расчет оптимальных значений рандомизированной стратегии связан с серьезными теоретическими и вычислительными трудностями. Без их преодоления остается только инженерный подход к расчету (p_1, \dots, p_N) (см. разд. 1). В последнем случае относительное преимущество диспетчеризации AA над RND достигает нескольких десятков процентов.

Установленное ранжирование стратегий сохраняется и при варьировании ценовых коэффициентов в формуле для одношаговых затрат. График зависимости предельных средних затрат L от ценового коэффициента α (при фиксированном коэффициенте $\beta = 1$ и загрузке 0,6) представлен на рис. 3. Значения параметров стратегий RND и AA при каждом значении α находились, как и в табл. 1, путем оптимизации на имитируемых траекториях.

Отметим, что при $\alpha = 0$ и $\beta = 1$, т. е. когда, по сути, предельные средние затраты связаны только с потерями поступающих заданий, наилучшая с точки зрения L стратегия Муорис оказываются хуже статической AA по критерию минимума предельного среднего времени T пребывания задания в системе¹. Впрочем, если отслеживается только один из рассматриваемых функционалов, то вне зависимости от загрузки системы динамическая стратегия Муорис, как и ожидалось²,

¹И наоборот: при $\alpha = 1$ и $\beta = 0$ наилучшая с точки зрения L динамическая стратегия Муорис оказывается хуже AA по критерию минимума предельной средней доли потерь, приходящейся на одно задание.

²Вследствие того, что очереди на серверах обслуживаются, по предположению, согласно дисциплине FIFO (first in, first out).

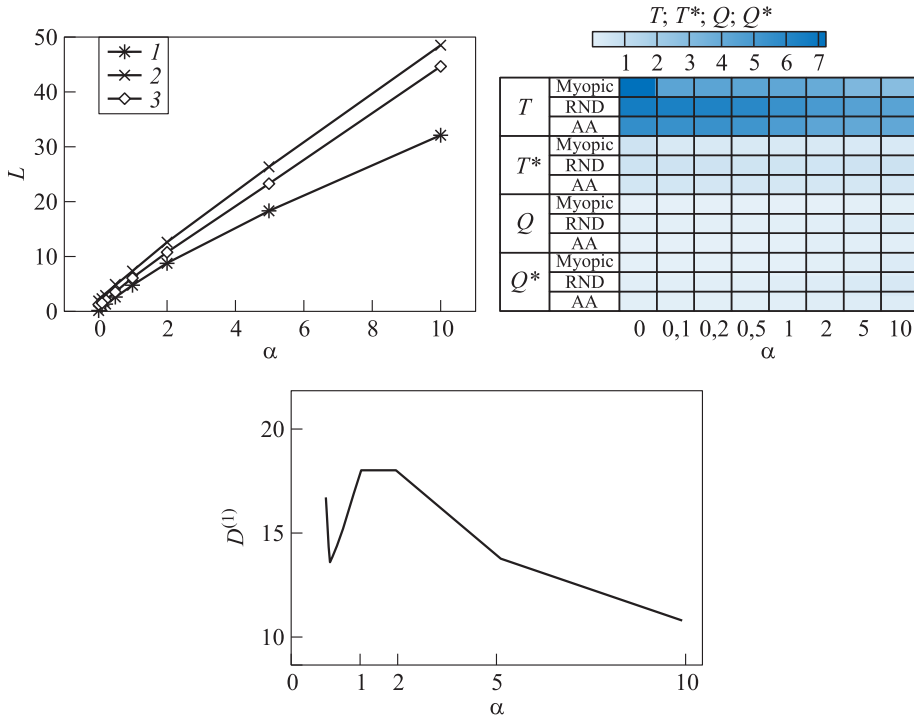


Рис. 3 Зависимость предельных средних затрат от ценового коэффициента α при $\lambda = 0,9$ и $\beta = 1$ для стратегий Муорис (1), RND (2) и AA (3)

остается предпочтительнее любой статической стратегии. Однако наблюдается следующий эффект: с уменьшением дисперсии объема задания эффективности стратегий AA и Муорис «сближаются», а стратегия RND продолжает сильно «отставать». Пусть, например, задания равномерного на $[10, 30]$ объема поступают в виде гиперэкспоненциального потока с параметрами $(0,9; 0,1)$ и $(0,1; 1)$ к диспетчеру, распределяющему их по $N = 100$ серверам со следующими характеристиками производительности и объема памяти:

$$R_i = 0,1 \left| 1 + 9 \left| \sin \left(\frac{0,02\pi}{i} \right) \right| \right|, \quad C_i = 5 + \left| 20 \left| \sin \left(5 + \frac{0,02\pi}{i} \right) \right| \right|, \quad 1 \leq i \leq 100.$$

Тогда все рассматриваемые функционалы принимают значения, приведенные в табл. 2 (предполагается, что $\alpha = 0,01$ и $\beta = 1$).

Как видно из табл. 2, диспетчеризация AA может практически не уступать в эффективности диспетчеризации, основанной на полных наблюдениях. Интуитивно оправданным кажется предположение, что наблюдения должны давать

Таблица 2 Значения всех функционалов для стратегий Муорис, RND и AA в системе из $N = 100$ неоднородных серверов, $D^{(1)} = 26$

Стратегия	T	T^*	Q	Q^*	L
Муорис	19,376	1	0,022	0,031	0,819
RND	26,653	1,447	0,064	0,079	1,852
AA	19,780	1,021	0,022	0,031	0,823

существенные преимущества. Однако оно подтверждается только для «сильно случайных» объемов заданий; в частности, для заданий фиксированного объема¹ алгоритмы, построенные по методу разд. 3 (в основном, при загрузке не выше средней²), могут выигрывать у диспетчеризаций по полным наблюдениям.

5 Заключение

Предложенный в разд. 3 метод дает параметрические стратегии, применимые при более общих, чем сделанные в разд. 2, предположениях о входящем потоке и дисциплинах обслуживания очередей. Так, ординарный входящий поток может быть заменен на групповой поток без внутренней структуры, а обслуживание в порядке поступления — на любую классическую консервативную дисциплину обслуживания (например, обслуживание задания наименьшего объема с прерыванием). Модифицируя соответствующим образом (2), новые диспетчеризации можно приспособить и к разным структурам системы; например, в случае многопроцессорных серверов рекурсию Линдли (2) необходимо заменить рекурсией Кифера–Вольфовица.

Как известно [3, 21], отличительная особенность систем добровольных вычислений — нестабильность работы (или, по-другому, частичная доступность) ресурсов. Поэтому в представленном в разд. 2 описании неявно предполагается, что речь идет о подмножестве серверов, которые демонстрируют стабильное поведение. В [22] на примере проекта добровольных вычислений `seti@home` продемонстрировано, что такое подмножество можно выделить. Путем варьирования значений параметров $D^{(1)}, \dots, D^{(K)}$ (и внесения, согласно [14], некоторых изменений в метод разд. 3) новые диспетчеризации, в отличие от статических, можно адаптировать к случаю, когда серверы нельзя считать абсолютно надежными, т. е. естественным образом отразить функциональные особенности исследуемой системы.

Наряду со всеми преимуществами новые диспетчеризации обладают и рядом недостатков. Главный из них — это отсутствие теоретической основы продук-

¹Примеры, когда это интуитивное соображение неверно, приведены, например, в [16].

²Отметим, что достигается преимущество путем более активного использования «медленных» серверов.

тивности столь простой в реализации, но интуитивно совершенно не очевидной конструкции. Нахождение условий, гарантирующих получение выигрыша от применения новых диспетчеризаций, представляется важным направлением дальнейших исследований. По имеющимся результатам вычислительных экспериментов подтверждается известный вывод о том, что получение выигрыша возможно, когда производительности серверов несильно отличаются друг от друга. Другое перспективное направление исследований — это разработка «аналитических» (в смысле [16]) алгоритмов диспетчеризации по предыстории для решения задач условной оптимизации (например¹, $T \rightarrow \min$ при известном ограничении на Q).

Литература

1. Коновалов М. Г., Разумчик Р. В. Комплексное управление в одном классе систем с параллельным обслуживанием // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 54–59. doi: 10.14357/19922264190409.
2. Соколов А. Е., Ушаков И. А. Математические методы моделирования при создании систем связи // Техника средств связи. Сер. АСУ, 1978. № 2. С. 3–11.
3. Mengistu T. M., Che D. Survey and taxonomy of volunteer computing // ACM Comput. Surv., 2019. Vol. 52. No. 3. Art. 59. 35 p. doi: 10.1145/3320073.
4. Nikitina N., Manzyuk M., Podlipnik Ć., Jukic M. Volunteer computing project SiDock@home for virtual drug screening against SARS-CoV-2 // Computer science protecting human society against epidemics / Eds. A. Byrski, T. Czachórski, E. Gelenbe, K. Grochla, Y. Murayama. — IFIP advances in information and communication technology book ser. — Cham: Springer, 2021. Vol. 616. P. 23–34. doi: 10.1007/978-3-030-86582-5_3.
5. Курочкин И. И., Григорьев К. В., Корлякова Е. Ю., Шилина А. М., Юрасова Е. П., Якимец В. Н. Исследование способов привлечения добровольцев в проекты добровольных распределенных вычислений // Информационное общество: образование, наука, культура и технологии будущего, 2022. Вып. 6. С. 49–61. doi: 10.17586/2587-8557-2022-6-49-61.
6. Lakshmi L. Resource and history-aware IoT task scheduling in volunteer assisted fog computing // Conference (International) on Smart Applications, Communications and Networking (SmartNets). — Piscataway, NJ, USA: IEEE, 2022. Art. 9993995. 6 p. doi: 10.1109/SmartNets55823.2022.9993995.
7. BOINC computing power. <https://boinc.berkeley.edu/computing.php>.
8. Мизин И. А., Кулешов А. П. Сети ЭВМ // Итоги науки и техники. Сер. Техническая кибернетика. — М.: ВИНТИ, 1986. Т. 20. С. 3–134.
9. Durrani M. N., Shamsi J. A. Review: Volunteer computing: Requirements, challenges, and solutions // J. Netw. Comput. Appl., 2014. Vol. 39. No. 1. P. 369–380. doi: 10.1016/j.jnca.2013.07.006.
10. Tapparello C., Funai C., Hijazi S., Aquino A., Karaoglu B., Ba H., Shi J., Heinzelman W. Volunteer computing on mobile devices: State of the art and future

¹Приближенное вычисление Q доступно благодаря известным эвристическим формулам для систем обслуживания конечной емкости (см., например, [23]).

- research directions // Enabling real-time mobile cloud computing through emerging technologies. — Hershey, PA, USA: IGI Global, 2015. P. 153–181. doi: 10.4018/978-1-4666-8662-5.ch005.
11. *Konovarov M., Razumchik R.* Using inter-arrival times for scheduling in non-observable queues // 31st European Conference on Modelling and Simulation Proceedings. — Budapest, Hungary: European Council for Modelling and Simulation, 2017. P. 667–672. doi: 10.7148/2017-0667.
 12. *Коновалов М. Г., Разумчик Р. В.* Обзор моделей и алгоритмов размещения заданий в системах с параллельным обслуживанием // Информатика и её применения, 2015. Т. 9. Вып. 4. С. 56–67. doi: 10.14357/1992264150406.
 13. *Hordijk A., van der Laan D. A.* Periodic routing to parallel queues and billiard sequences // Math. Method. Oper. Res., 2004. Vol. 59. No. 2. P. 173–192. doi: 10.1007/s001860300322.
 14. *Konovarov M., Razumchik R.* Minimizing mean response time in batch-arrival non-observable systems with single-server FIFO queues operating in parallel // Communications ECMS, 2021. Vol. 35. No. 1. P. 272–278. doi: 10.7148/2021-0272.
 15. *Коновалов М. Г., Разумчик Р. В.* О размещении заданий на двух серверах при неполном наблюдении // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 57–67. doi: 10.14357/1992264160406.
 16. *Konovarov M., Razumchik R.* Improving routing decisions in parallel non-observable queues // Computing, 2018. Vol. 100. No. 10. P. 1059–1079. doi: 10.1007/s00607-018-0598-5.
 17. *Semchedine F., Bouallouche-Medjkoune L., Aissani D.* Task assignment policies in distributed server systems: A survey // J. Netw. Comput. Appl., 2011. Vol. 34. No. 4. P. 1123–1130. doi: 10.1016/j.jnca.2011.01.011.
 18. *Hordijk W., Hordijk A., Heidergott B.* A genetic algorithm for finding good balanced sequences in a customer assignment problem with no state information // Asia Pac. J. Oper. Res., 2015. Vol. 32. No. 3. Art. 1550015. 20 p. doi: 10.1142/S0217595915500153.
 19. *Назин А. В., Позняк А. С.* Адаптивный выбор вариантов: рекуррентные алгоритмы. — М.: Наука, 1986. 288 с.
 20. *Коновалов М. Г.* Методы адаптивной обработки информации и их приложения. — М.: ИПИ РАН, 2007. 212 с.
 21. *Заикин О. С., Посыпкин М. А., Семёнов А. А., Храпов Н. П.* Опыт организации добровольных вычислений на примере проектов optima@home и sat@home // Вестник Нижегородского университета им. Н. И. Лобачевского, 2012. № 5-2. С. 340–347. EDN: PZKQIZ.
 22. *Javadi B., Kondo D., Vincent J. M., Anderson D. P.* Discovering statistical models of availability in large distributed systems: An empirical study of seti@home // IEEE T. Parall. Distr., 2011. Vol. 22. No. 11. P. 1896–1903. doi: 10.1109/TPDS.2011.50.
 23. *Tijms H.* Heuristics for finite-buffer queues // Probab. Eng. Inform. Sc., 1992. Vol. 6. No. 3. P. 277–285. doi: 10.1017/S0269964800002540.

Поступила в редакцию 09.12.22

DISPATCHING IN NONOBSERVABLE PARALLEL QUEUES WITH FINITE CAPACITIES

M. G. Konovalov and R. V. Razumchik

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Consideration is given to the problem of optimal centralized routing in one simple model of volunteer computer systems. The model consists of the finite number of parallel finite-capacity FIFO (first in, first out) queues, each with a single server, and a single dispatcher. Homogeneous jobs arrive one-by-one in a stochastic manner to the dispatcher which must instantly make a routing decision based solely on its previous routing decisions and, possibly, the time instants at which those decisions were made. Although no feedback from the queues is available, the dispatcher knows the maximum queues’ capacities, all service rates, the distribution of job sizes, and the inter-arrival time distribution. The new algorithm for generating routing decisions “on the fly” which optimizes the given cost function of the stationary mean sojourn time and the loss probability is proposed.

Keywords: parallel service systems; dispatching; resource allocation policies; control under incomplete observations; program control

DOI: 10.14357/08696527230303

EDN: XUVVNH

Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

References

1. Konovalov, M., and R. Razumchik. 2019. Kompleksnoe upravlenie v odnom klasse sistem s parallel’nym obsluzhivaniem [Mixed policies for online job allocation in one class of systems with parallel service]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):54–59. doi: 10.14357/19922264190409.
2. Sokolov, A. E., and I. A. Ushakov. 1978. Matematicheskie metody modelirovaniya pri sozdanii sistem svyazi [Mathematical modeling methods for creating communication systems]. *Tekhnika sredstv svyazi. Ser. ASU* [Communication Technology. Automatic Control Systems Series] 2:3–11.
3. Mengistu, T. M., and D. Che. 2019. Survey and taxonomy of volunteer computing. *ACM Comput. Surv.* 52(3):59. 35 p. doi: 10.1145/3320073.
4. Nikitina, N., M. Manzyuk, Č. Podlipnik, and M. Jukic. 2021. Volunteer computing project SiDock@home for virtual drug screening against SARS-CoV-2. *Computer*

- science protecting human society against epidemics. Eds. A. Byrski, T. Czachórski, E. Gelenbe, K. Grochla, and Y. Murayama. IFIP advances in information and communication technology book ser. Cham: Springer. 616:23–34. doi: 10.1007/978-3-030-86582-5_3.
5. Kurochkin, I. I., K. V. Grigoriev, E. U. Korlyakova, A. M. Shilina, E. P. Urasova, and V. N. Yakimets. 2022. Issledovanie sposobov privlecheniya dobrovol'tsev v proekty dobrovol'nykh raspredelennykh vychisleniy [Study of methods to attract volunteers in voluntary distributed computing projects]. *Informatsionnoye obshchestvo: obrazovanie, nauka, kul'tura i tekhnologii budushchego* [The information society: Education, science, culture, and future technologies] 6:49–61. doi: 10.17586/2587-8557-2022-6-49-61.
 6. Lakshmi, L. 2022. Resource and history-aware IoT task scheduling in volunteer assisted fog computing. *Conference (International) on Smart Applications, Communications and Networking*. Piscataway, NJ: IEEE. 9993995. 6 p. doi: 10.1109/SmartNets55823.2022.9993995.
 7. BOINC computing power. Available at: <https://boinc.berkeley.edu/computing.php> (accessed September 20, 2023).
 8. Mizin, I. A., and A. P. Kuleshov. 1986. *Seti EVM* [Computer networks]. *Itogi nauki i tekhniki. Ser. Tekhnicheskaya kibernetika* [Technical cybernetics. Results of science and technology ser]. Moscow: VINITI. 20:3–134.
 9. Durrani, M. N., and J. A. Shamsi. 2014. Volunteer computing: Requirements, challenges, and solutions. *J. Netw. Comput. Appl.* 39(1):369–380. doi: 10.1016/j.jnca.2013.07.006.
 10. Tapparello, C., C. Funai, S. Hijazi, A. Aquino, B. Karaoglu, H. Ba, J. Shi, and W. Heinzelman. 2015. Volunteer computing on mobile devices: State of the art and future research directions. *Enabling real-time mobile cloud computing through emerging technologies*. Hershey, PA: IGI Global. 153–181. doi: 10.4018/978-1-4666-8662-5.ch005.
 11. Konovalov, M., and R. Razumchik. 2017. Using inter-arrival times for scheduling in non-observable queues. *31st ECMS Conference (International) on Modelling and Simulation Proceedings*. Budapest, Hungary: ECMS. 667–672. doi: 10.7148/2017-0667.
 12. Konovalov, M. G., and R. V. Razumchik. 2015. Obzor modeley i algoritmov razmeshcheniya zadaniy v sistemakh s parallel'nym obsluzhivaniem [Methods and algorithms for job scheduling in systems with parallel service: A survey]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):56–67. doi: 10.14357/1992264150406.
 13. Hordijk, A., and D. A. van der Laan. 2004. Periodic routing to parallel queues and billiard sequences. *Math. Method. Oper. Res.* 59(2):173–192. doi: 10.1007/s001860300322.
 14. Konovalov, M., and R. Razumchik. 2021. Minimizing mean response time in batch-arrival non-observable systems with single-server FIFO queues operating in parallel. *Communications ECMS* 35(1):272–278. doi: 10.7148/2021-0272.
 15. Konovalov, M., and R. Razumchik. 2016. O razmeshchenii zadaniy na dvukh serverakh pri nepolnom nablyudenii [Dispatching to two parallel nonobservable queues using only static information]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):57–67. doi: 10.14357/1992264160406.
 16. Konovalov, M., and R. Razumchik. 2018. Improving routing decisions in parallel non-observable queues. *Computing* 100(10):1059–1079. doi: 10.1007/s00607-018-0598-5.

17. Semchedine, F., L. Bouallouche-Medjkoune, and D. Aissani. 2011. Task assignment policies in distributed server systems: A survey. *J. Netw. Comput. Appl.* 34(4):1123–1130. doi: 10.1016/j.jnca.2011.01.011.
18. Hordijk, W., A. Hordijk, and B. Heidergott. 2015. A genetic algorithm for finding good balanced sequences in a customer assignment problem with no state information. *Asia Pac. J. Oper. Res.* 32(3):1550015. 20 p. doi: 10.1142/S0217595915500153.
19. Nazin, A. V., and A. S. Poznyak. 1986. *Adaptivnyy vybor variantov: rekurrentnye algoritmy* [Adaptive option selection: Recursive algorithms]. Moscow: Nauka. 288 p.
20. Kononov, M. G. 2007. *Metody adaptivnoy obrabotki informatsii i ikh prilozheniya* [Methods of adaptive information processing and their applications]. Moscow: IPI RAN. 212 p.
21. Zaikin, O. S., M. A. Posypkin, A. A. Semenov, and N. P. Khrapov. 2012. Opyt organizatsii dobrovol'nykh vychisleniy na primere proektov optima@home i sat@home [Experience in organizing voluntary computing on the example of the optima@home and sat@home projects]. *Vestnik Nizhegorodskogo universiteta im. N. I. Lobachevskogo* [Bulletin of the Lobachevsky University of Nizhni Novgorod] 5-2:340–347. EDN: PZKQIZ.
22. Javadi, B., D. Kondo, J.-M. Vincent, and D. P. Anderson. 2011. Discovering statistical models of availability in large distributed systems: An empirical study of seti@home. *IEEE T. Parallel. Distr.* 22(11):1896–1903. doi: 10.1109/TPDS.2011.50.
23. Tijms, H. 1992. Heuristics for finite-buffer queues. *Probab. Eng. Inform. Sc.* 6(3):277–285. doi: 10.1017/S0269964800002540.

Received December 9, 2022

Contributors

Kononov Mikhail G. (b. 1950) — Doctor of Science in technology, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkononov@ipiran.ru

Razumchik Rostislav V. (b. 1984) — Doctor of Science in physics and mathematics, deputy director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; rrazumchik@ipiran.ru

ПРИНЦИПЫ РАЗРАБОТКИ ПРОГРАММНОГО КОМПЛЕКСА ЦЕНТРА УПРАВЛЕНИЯ НАЦИОНАЛЬНОЙ ИССЛЕДОВАТЕЛЬСКОЙ КОМПЬЮТЕРНОЙ СЕТЬЮ РОССИИ*

А. Г. Абрамов¹, В. А. Порхачёв²

Аннотация: Статья посвящена общим принципам и методикам разработки и совершенствования программного комплекса цифрового центра управления национальной исследовательской компьютерной сетью нового поколения (ЦУС НИКС). В работе исследованы и систематизированы задачи, решаемые центрами управления крупными телекоммуникационными сетями, включая и специфичные для национальных научно-образовательных сетей. По результатам выделены основные задачи и функции ЦУС НИКС, приведена аргументация и обоснование создания собственного программного решения, обозначена его новизна, отличительные конкурентные преимущества и уникальные функции, предложены возможные направления развития и практического использования. Специальное внимание уделено представлению и обсуждению особенностей архитектуры программного ядра комплекса, а также взаимосвязанных базовых и дополнительных компонентов, включая набор функциональных и вспомогательных модулей в составе ядра, микросервисов для взаимодействия с внешними системами и внешних сервисов.

Ключевые слова: национальная исследовательская компьютерная сеть нового поколения, НИКС; центр управления сетью, ЦУС; мониторинг и управление сетями; программный комплекс

DOI: 10.14357/08696527230304

EDN: PRLQXO

1 Введение

Научно-технологическая задача мониторинга и управления крупными телекоммуникационными инфраструктурами на протяжении многих лет привлекает к себе повышенное внимание исследователей и инженеров. На сегодняшний день проработаны теоретические основы, предложены методики, практические приемы и технические средства, создан большой объем специализированного программного обеспечения для решения широкого спектра функциональных задач в рамках процессов мониторинга и управления, активно

*Статья подготовлена в рамках государственного задания ФГУ ФНЦ НИИСИ РАН по теме № FNEF-2022-0014. Работа выполнена с использованием ресурсов Центра коллективного пользования «Национальная исследовательская компьютерная сеть».

¹Санкт-Петербургское отделение Межведомственного суперкомпьютерного центра Российской академии наук — филиала ФГУ ФНЦ НИИСИ РАН, abramov@niks.su

²Санкт-Петербургское отделение Межведомственного суперкомпьютерного центра Российской академии наук — филиала ФГУ ФНЦ НИИСИ РАН, porhachev@niks.su

привлекаются методы искусственного интеллекта и машинного обучения (см., к примеру, монографии [1–3] и ссылки в них).

Постоянно расширяющаяся пользовательская база, интенсивность использования массово создаваемых цифровых сервисов и требования к качеству их предоставления, лавинообразный рост объемов сетевого трафика, широкий спектр сетевых архитектур и технологий, различного используемого на практике сетевого оборудования и программного обеспечения, растущая интенсивность и разнообразие угроз кибербезопасности заставляют специалистов продолжать совершенствовать существующие и разрабатывать новые методы и инструменты, направленные на обеспечение эффективного, надежного и бесперебойного функционирования сетевых инфраструктур и сервисов, предоставляемых на их основе.

Национальные научно-образовательные сети (National Research and Education Networks, NREN) обладают выраженными особенностями, отличающими их от коммерческих операторов сетей общего пользования [4–7], во взаимной кооперации такие сети уделяют большое внимание разработке и совершенствованию научных и методических основ, экспертных и практических рекомендаций, программных инструментов для мониторинга и управления сетями, отражающих текущее состояние отрасли информационно-коммуникационных (ИКТ) в целом и ориентированных на использование в совместной работе интегрированной инфраструктуры мировых NREN в интересах глобального научно-образовательного сообщества и иных заинтересованных пользователей.

Национальная исследовательская компьютерная сеть нового поколения (НИКС, <https://niks.su>), созданная в 2019 г. и функционирующая в качестве NREN России [7, 8], наследующая накопленный опыт и наработки университетской сети RUNNet и академической сети RASNet, в настоящее время вовлечена в систематическое изучение, анализ и внедрение лучших практик ведущих NREN с их адаптацией к внутрироссийским особенностям и с учетом утвержденного в рамках национального проекта «Наука и университеты» плана развития сети на период 2021–2024 гг.

Осуществление в постоянном режиме надежного и объективного мониторинга и контроля функционирования таких специализированных сетей в качестве одного из ключевых компонентов более широкого класса задач управления сетью имеет крайне важное значение, в том числе с учетом нацеленности на удовлетворение повышенных требований целевой группы пользователей к качеству обслуживания при участии в крупных проектах класса «мегасайенс» и иных значимых научно-образовательных и научно-технических проектах и инициативах, реализуемых в составе коллабораций и интенсивно использующих современные средства ИКТ.

Хорошо известный специалистам термин Network Operations Center (NOC, или в русскоязычной транскрипции — центр управления сетью, ЦУС) означает централизованную локацию, используемую сетевыми администраторами для целей непрерывного отслеживания и контроля работоспособности сети. Ключе-

вые функции ЦУС всегда включают в себя мониторинг и ряд тесно связанных операций, таких как инвентаризация, конфигурация, управление ресурсами, статистика и отчетность и т. д. [9, 10].

Центры управления сетями для решения закрепленных за ними задач используют специализированные программные комплексы и системы, которые могут обеспечивать только одну или сразу несколько функций, различаться способами развертывания (локальные, облачные), распространяться на коммерческих условиях или на основе открытых лицензий [9–11].

Одной из целей настоящей работы ставилось проведение комплексного исследования по выявлению и систематизации задач, традиционно решаемых ЦУС крупных операторов связи, включая и задачи, выходящие за рамки их деятельности и специфичные для отраслевых сетей класса NREN. Результатом исследований и анализа некоторых из существующих программных решений стал обоснованный выбор для проектирования и реализации набора взаимосвязанных модулей в составе программного комплекса цифрового ЦУС НИКС, функции и основные возможности которых перечислены в работе.

В статье также рассмотрены базовые принципы и методики разработки комплекса, представлены и обсуждены особенности верхнеуровневой архитектуры программного ядра, функциональные и вспомогательные модули и компоненты в его составе, специальные службы и микросервисы для взаимодействия с задействованными для решения отдельных задач внешними системами и внешние сервисы.

2 Цели, задачи и основные функции программного комплекса в составе цифрового ЦУС НИКС

Функционировавшее ранее и выведенное из эксплуатации в 2021 г. программное решение ЦУС НИКС было разработано еще в середине 1990-х гг. Оно не удовлетворяло актуальным стандартам сферы ИКТ, требованиям информационной безопасности, эргономичности пользовательских интерфейсов, базировалось на устаревшем программном обеспечении, отсутствовали возможности масштабирования, полноценной интеграции с внешними системами, расширенной статистики и аналитики данных.

Продолжение эксплуатации этой версии решения с учетом реализации планов ускоренного развития НИКС в 2021–2024 гг. как важного отраслевого компонента ИКТ-инфраструктуры страны и кратного увеличения числа пользователей могло повлечь за собой существенные риски, связанные с возможным частичным или полным отказом работоспособности и необеспечением критичных функций мониторинга, подверженности различного рода сетевым угрозам, снижением эффективности функционирования службы технической поддержки.

Обзор тематической литературы, описанных в открытых источниках коммерческих реализаций ЦУС, апробация ряда свободно распространяемых продуктов, учет экспертного опыта ведущих NREN позволили сформировать набор

решаемых подобными системами задач — мониторинг сети и сервисов, управление проблемами, обеспечение жизненного цикла заявок клиентов и проблемных билетов, управление конфигурациями, управление знаниями и документацией, мониторинг и управление производительностью, управление инвентаризацией и учет ресурсов, статистика и отчетность, агрегация, визуализация и представление данных, управление качеством предоставляемых услуг, обеспечение информационной безопасности [9, 10]. Дополнительно могут ставиться и решаться задачи, связанные с отраслевой спецификой NREN, например учета сетевого взаимодействия между пользователями, статистического анализа в разрезе административно-территориальных единиц, типов организаций и т. п.

Массовый уход из страны зарубежных цифровых компаний, в том числе ведущих разработчиков систем мониторинга и управления крупными телекоммуникационными инфраструктурами (SolarWinds, ManageEngine, Paessler PRTG, CA Spectrum, IBM Tivoli Netcool, ServiceNow, Microsoft SCOM и др.), прекращение продаж и технической поддержки, несоответствие на текущем этапе развития цены и качества российских коммерческих продуктов, трудно прогнозируемое время внедрения с учетом имеющихся требований оставили возможности для применения свободно распространяемых пакетов с открытым исходным кодом или разработки нового программного решения.

Комплексное исследование функциональных возможностей и интерфейсов свободно распространяемых продуктов как кандидатов для использования в качестве базового решения, отдельных компонентов или внешних по отношению к ЦУС НИКС систем было проведено для таких пакетов, как NOC Project, GLPI, Zabbix, Grafana/Prometheus, Nagios Core, OTRS, Cacti, ELK Stack, RackTables, NetBox и phpIPAM.

Большинство из перечисленных продуктов ориентировано на решение какой-то одной или только нескольких смежных задач в рамках типичного для систем класса ЦУС набора функций и операций. Такая ограниченность функционала привела бы к необходимости внедрения сразу нескольких решений для покрытия сформированных для цифрового ЦУС НИКС функционально-технических требований, разработки средств интеграции («коннекторов», API) для обеспечения их взаимодействия, внесения изменений в структуру баз данных и в соответствующие методы обработки для представления организаций с расширенным набором атрибутов (см. подробнее ниже).

В этих условиях и с учетом устойчивого курса на импортозамещение в сфере ИКТ было принято решение о разработке новой версии программного комплекса, способного обеспечить единую интегрированную среду эффективного организационного и технологического сопровождения работ по обеспечению текущего функционирования и развития федеральной отраслевой сети НИКС с акцентом на следующие верхнеуровневые задачи [11]:

- создание отказоустойчивого, высокопроизводительного интегрированного программного комплекса ЦУС НИКС в соответствии с подлежащими автоматизации функциями;

- обеспечение единой точки аутентификации и авторизации и гибкого ролевого доступа к функциям и интерфейсам;
- сбор, обработка и централизованное представление данных о функционировании сети и сервисов НИКС;
- создание и поддержание в актуальном состоянии единой базы данных по организациям сферы науки и высшего образования (действующие и потенциальные пользователи), а также коммерческим организациям-контрагентам по проекту НИКС.

Программный комплекс ориентирован на задействие лучших практик систем подобного рода, предназначен для автоматизации процессов управления НИКС, обеспечения централизованного сбора и агрегации данных о функционировании сети и сервисов, текстового и визуального представления сетевым инженерам и представителям потребителей информации об использовании сетевых ресурсов и сервисов, предоставления широкого набора инструментов интегрального мониторинга и планирования развития сети и сервисов для руководства проекта.

Основными целями создания комплекса ставились:

- разработка автоматизированного решения ЦУС НИКС с развитыми интегрированными возможностями мониторинга и управления сетью и предоставляемыми сервисами;
- реализация возможностей расширенного статистического и аналитического учета функционирования сети и сервисов;
- повышение качества представления информации об использовании организациями сетевых ресурсов и сервисов НИКС;
- повышение эффективности функционирования службы технической поддержки, сокращение времени выявления сетевых инцидентов.

Новизна комплекса в сравнении с конкурентными продуктами заключается в широком покрытии типовых задач мониторинга и управления, включая взаимоотношения с пользователями и поставщиками услуг, информационно-технологические (ИТ) активы, обеспечивающие документы (договоры, бланки заказа, соглашения с целевыми пользователями) и другие компоненты процесса. Взаимная увязанность объектов в базе данных, наличие сведений как технического, так и финансово-экономического характера, интеграции с внешними системами и сервисами сбора метрик мониторинга, статистики и аналитики сетевого трафика, журналов устройств предоставляют возможности для формирования комплексных отчетов и аналитических срезов, обоснованной формулировки стратегических и оперативных направлений развития сети и сервисов, вносят вклад в повышение общей эффективности решения задач.

Доступность исходного кода и следование основным принципам современной методологии гибкой разработки программного обеспечения (Agile) позволяют оперативно исправлять выявляемые ошибки и недочеты, добавлять новые

функции и улучшения в веб-интерфейсе, реагируя на пожелания пользователей, проектировать и разрабатывать новые функциональные модули и программные компоненты.

Уникальные функции и возможности комплекса связаны с наличием в базе данных актуальной информации обо всех образовательных организациях высшего образования и науки, находящихся в подчинении разных федеральных органов исполнительной власти, а также об отдельных государственных и коммерческих организациях, участвующих в реализации приоритетных направлений научно-технологического развития, определенных в соответствии со Стратегией научно-технологического развития Российской Федерации.

Накопленная информация помимо официальных сведений, контактных персон, общего описания и статистики потребляемых услуг, договорных отношений (для контрагентов) включает сведения об участии организаций в ключевых проектах и подпрограммах государственной программы «Научно-технологическое развитие Российской Федерации» и национального проекта «Наука и университеты», в том числе в проектах по созданию университетских кампусов мирового уровня, по созданию и развитию научно-образовательных и научных центров мирового уровня, центров компетенций национальной технологической инициативы, центров коллективного пользования научным оборудованием и уникальных научных установок, в проекте «Передовые инженерные школы», в программе «Приоритет-2030» и др.

Кроме того, в базу данных включены сведения об определенных национальным проектом «Наука и университеты» критериях [7, 8], на основании которых осуществляется отбор ведущих организаций для подключения к НИКС (в частности, принадлежность к первой категории научных организаций, ведомственная подчиненность Правительству Российской Федерации, статус федерального или национального исследовательского университета).

Агрегированная информация используется для формирования планов территориального развития НИКС, планирования подключения к сети новых организаций, может послужить основой для создания специализированных отраслевых платформ и сервисов, ориентированных на централизованный учет и предоставление возможностей для комплексной оценки научного и образовательного потенциала организаций и субъектов.

3 Особенности архитектуры программного комплекса цифрового ЦУС НИКС

Архитектура программного комплекса цифрового ЦУС НИКС проектировалась на основе системного подхода с выделением центральной части (монолитного ядра) с набором тесно взаимодействующих и функционально дополняющих друг друга внутренних модулей и служб, а также микросервисных компонентов, отвечающих за обмен данными с внешними системами и сервисами (рис. 1) [11]. Комплекс представляет собой структурно сложное динамическое веб-приложе-

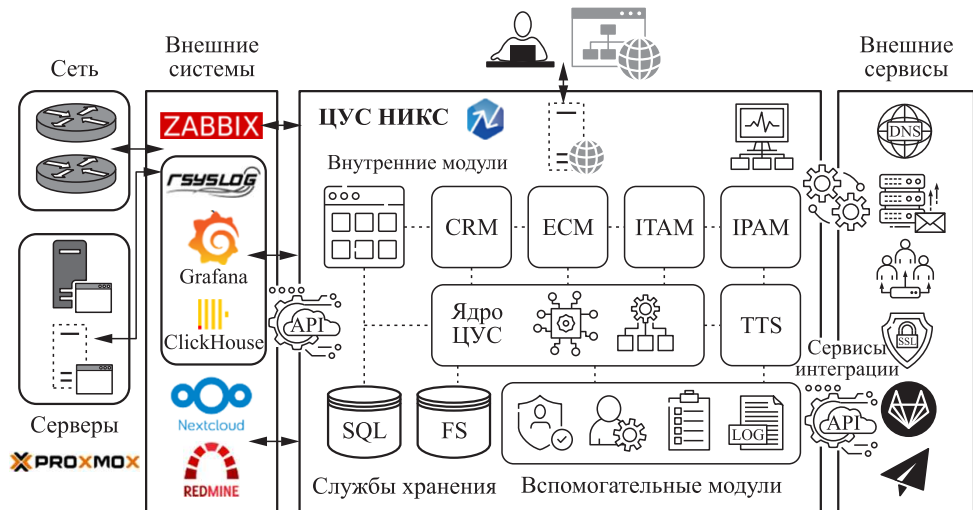


Рис. 1 Архитектурная схема программного комплекса и структурных компонентов в составе цифрового ЦУС НИКС

ние, оперирующее множественными потоками данных и связанное с внешним программным окружением.

Модули, службы и сервисы комплекса (с учетом внешнего окружения) можно классифицировать в составе следующих основных групп:

- *внутренние модули* (набор основных функциональных и вспомогательных модулей в составе ядра комплекса);
- *службы хранения* (базы данных MariaDB, PostgreSQL, ClickHouse и обычные файловые хранилища);
- *сервисы интеграции* (микросервисы для взаимодействия с внешними системами и сервисами через HTTP API — Zabbix, Grafana, NextCloud, а также решение HelpDesk/Trouble Ticket, TTS, собственной разработки);
- *внешние сервисы* (набор внешних взаимодействующих с комплексом сервисов и приложений — Git, DNS, служба электронной почты, API Telegram, сервис цифровых SSL-сертификатов и др.).

Обозначенный в предыдущем разделе богатый спектр задач, которые должны обеспечиваться развитыми решениями ЦУС, обусловил необходимость выбора на первом этапе перечисленного далее относительно ограниченного набора программных компонентов в качестве приоритетных для разработки.

Один из выводов по результатам проведенных исследований и накопленного опыта состоит в том, что к решению отдельных задач, требующих существенных

затрат при самостоятельной разработке соответствующих программных модулей и подсистем с нуля, целесообразно привлекать функционально развитые, прошедшие апробацию и широко внедренные в отраслевую практику свободно распространяемые продукты. В случае с ЦУС НИКС задачи первичного мониторинга сети и сервисов, управления проблемами и документацией, визуализации и представления данных, статистики и отчетности решаются с задействованием функционала специализированных пакетов Zabbix, Grafana, Logstash и NextCloud.

Такой подход представляется весьма эффективным и экономичным в отношении затрат и сроков внедрения и может быть рекомендован организациям, которые ориентированы на собственные разработки и/или использование свободно распространяемых решений для выполнения мониторинга и управления цифровыми инфраструктурами.

Подлежащий расширению по мере поэтапного развития цифрового ЦУС НИКС набор внутренних функциональных модулей на текущий момент включает в себя следующие:

- модуль управления взаимоотношениями с контрагентами (Customer Relationship Management, CRM);
- модуль управления корпоративным контентом и знаниями (Enterprise Content Management, ECM);
- модуль управления ИТ-активами (IT Asset Management, ITAM);
- модуль управления IP-адресами и подсетями (IP Address Management, IPAM).

Модуль CRM («Цифровой профиль организации») отвечает за управление сведениями о контрагентах и контактных персонах, централизованный учет и представление информации об использовании сетевых ресурсов и сервисов НИКС, в том числе о подключениях и сетевом взаимодействии (оборудование,

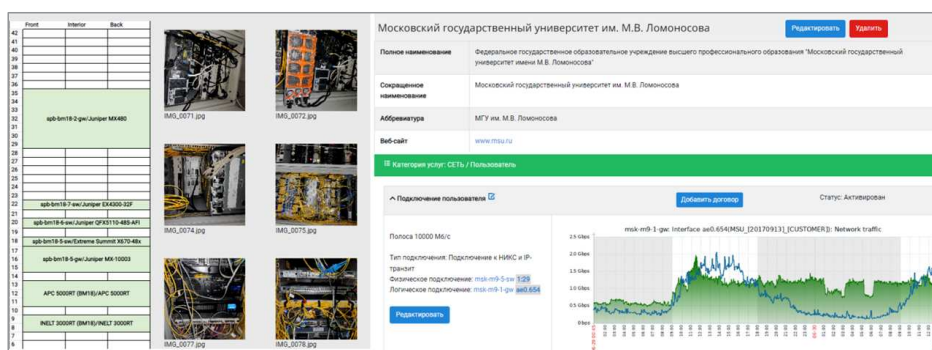


Рис. 2 Примеры веб-интерфейсов программного комплекса

порты, тип подключения, пропускная способность), а также за графики загрузки портов (рис. 2). Информация о подключениях импортируется из модуля ИТАМ, а графики и статистика сетевого взаимодействия — из системы мониторинга Zabbix. Функционал модуля позволяет управлять сервисами предустановленных типов (для пользователей и провайдеров услуг), добавлять к ним атрибуты (или автоматически импортировать их из Zabbix), «прикреплять» к сервисам договоры (связь с модулем ЕСМ) и другую востребованную информацию.

Модулем ЕСМ обеспечиваются функции работы с документами (доходные и расходные договоры, дополнительные соглашения, бланки заказов, счета-фактуры и др.), включая просмотр списков документов и управление ими, привязку документа к контрагенту и услуге, расчет доходов (расходов) за интересующие периоды времени и с учетом типов услуг и др. Файлы документов размещаются в формате PDF в корпоративной веб-системе для совместной работы NextCloud с указанием в атрибутах гиперссылки для возможности просмотра с помощью специального плагина.

Модуль ИТАМ предназначен для централизованного учета и управления ИТ-активами НИКС, включая узлы, стойки, устройства (сетевые, серверные, технологические), модули, каналы связи, виртуальные и иные ресурсы, задействованные в работе сети и сервисов. Модуль выполняет функции ведения реестров объектов, учета оборудования с привязкой к локациям хранения/размещения и представления в наглядном виде стоек с оборудованием, поддержки жизненного цикла, динамической визуализации карты сети, каналов связи с отображением их загрузки и состояния. Оперативная информация об устройствах, интерфейсах, метриках и состояниях выгружается из Zabbix.

Модуль ИРАМ отвечает за централизованное планирование и учет подсетей и IPv4/IPv6-адресов, отображение иерархического списка подсетей с показом свободного адресного пространства, предоставление функционала поиска адресов, управления подсетями и адресами, создания отчетов. Задачи начального обнаружения активных IP-адресов в заданных подсетях, периодического автоматического сканирования подсетей с проверкой состояния адресов, проактивного мониторинга и оповещения о проблемах, связывания адресов с устройствами и интерфейсами решаются модулем ИТАМ с отображением в интерфейсе ИРАМ.

В круг вспомогательных входят следующие модули комплекса с соответствующим предназначением:

- *авторизации и управления доступом* (регистрация, аутентификация, управление учетными записями пользователей, реализация гибкой ролевой модели доступа);
- *справочников* (ведение справочников, в том числе географических, адресного с визуализацией на карте, моделей оборудования и др.);
- *журналирования* (хранение и отображение сведений о системных событиях, действиях конечных пользователей);

- *интеграционного взаимодействия* (реализация централизованного и унифицированного событийно-ориентированного обмена данными с внешними системами).

Направления развития комплекса и связанных систем предполагают как разработку новых, так и дальнейшее совершенствование существующих основных функциональных и вспомогательных модулей. Функционал будет расширен, в частности, за счет модуля отчетности, который позволит автоматически формировать структурированные текстовые и графические отчеты, консолидирующие данные из различных источников, компоновать информационные панели («дашборды») для разных категорий пользователей, а также коллекции виджетов с возможностью встраивания в веб-интерфейс комплекса и экспорта во внешние системы.

Важной задачей представляется внедрение технологии единого входа (Single Sign-On, SSO) с аутентификацией через корпоративный сервер LDAP (Keycloak), а также разработка собственного API. В планах также значится развитие существующих и реализация новых интеграций с внедряемыми и систематически совершенствуемыми высоконагруженными сервисами расширенной статистики и аналитики сетевого трафика, журналов сетевых устройств и серверов, решениями информационной безопасности в целях обогащения представляемой информации дополнительными данными мониторинга, статистической и аналитической информацией, выполнения интеллектуальной корреляции событий и выявления инцидентов.

Заметим в завершение, что программный стек ядра комплекса базируется на стандартных для отрасли веб-разработки языках PHP («бэкенд», фреймворк веб-приложений Laravel) и JavaScript («фронтенд», библиотека jQuery).

4 Заключение

Надежное и эффективное решение задач мониторинга и управления глобальными телекоммуникационными сетями и предоставляемыми на их основе сервисами является необходимым условием успешной эксплуатации и реализации планов развития таких сетей. Центр управления сетью НИКС в ходе своей модернизации стремится соответствовать современным достижениям и ключевым тенденциям отрасли, а также принимать во внимание особенности, связанные с целевой направленностью сети и предоставляемых сервисов.

Специфичные задачи и функции систем мониторинга и управления национальными научно-образовательными сетями предъявляют особые требования к реализациям соответствующих программных решений.

По результатам анализа возможностей использования и функционала нескольких в той или иной степени вошедших в практику решений авторами был сделан выбор в пользу разработки собственного программного комплекса цифрового ЦУС НИКС с ядром в виде непрерывно развиваемой системы и окружения

в составе специально отобранных, протестированных и внедренных пакетов с открытым исходным кодом.

Совмещение и взаимное обогащение функциональных компонентов комплекса в рамках системного подхода показало на первых этапах эксплуатации свою высокую эффективность, а по целому ряду функций и возможностей в контексте перспектив развития продукта можно ожидать характеристик, сопоставимых или лучших по сравнению с ведущими зарубежными коммерческими и доступными свободно распространяемыми продуктами.

Литература

1. *Claise B., Wolter R.* Network management: Accounting and performance strategies. — Cisco Press, 2012. 631 p.
2. *Subramanian M., Gonsalves T. A., Usha Rani N.* Network management: Principles and practice. — 2nd ed. — Prentice Hall, 2012. 695 p.
3. *Ge J., Li T., Wu Y.* AI and machine learning for network and security management. — IEEE Press ser. on networks and service management. — Wiley/IEEE Press, 2022. 304 p.
4. *McClure C. R., Bishop A. P., Doty P., Rosenbaum H.* The national research and education network (NREN): Research and policy perspectives. — Praeger, 1991. 758 p.
5. *Allocchio C., Balint L., Berkhout V., et al.* A history of international research networking: The people who made it happen. — New York, NY, USA: Wiley/VCH, 2010. 317 p.
6. GÉANT: The case for NRENs. A repository of resources to support funding, advocacy and the advancement of national and regional R&E networks. <https://caseforrens.org>.
7. *Абрамов А. Г., Гончар А. А., Евсеев А. В., Шабанов Б. М.* Национальная исследовательская компьютерная сеть нового поколения: текущее состояние и концепция развития // Информационные технологии, 2021. Т. 27. № 3. С. 115–124. doi: 10.17587/it.27.115-124.
8. *Абрамов А. Г., Евсеев А. В., Гончар А. А., Шабанов Б. М.* Вопросы увеличения пропускной способности и территориальной доступности национальной исследовательской компьютерной сети России // Системы и средства информатики, 2022. Т. 32. № 2. С. 4–12. doi: 10.14357/08696527220201.
9. SIG-NOC Tools Survey: What software tools R&E operations centres use. <https://wiki.geant.org/display/SIGNOC/SIG-NOC+Tools+Survey+2019>.
10. *Abramov A. G.* Issues of modernization of the monitoring and control system of the National Research Computer Network of Russia with an emphasis on free software solutions // Lobachevskii J. Mathematics, 2021. Vol. 42. Iss. 11. P. 2469–2480. doi: 10.1134/S1995080221110020.
11. *Abramov A. G., Porkhachev V. A.* Architecture and main components of the improved digital center for monitoring and management of the National Research Computer Network of Russia // Lobachevskii J. Mathematics, 2022. Vol. 43. Iss. 10. P. 2822–2832. doi: 10.1134/S1995080222130029.

Поступила в редакцию 21.07.23

PRINCIPLES OF DEVELOPMENT OF A SOFTWARE SUITE FOR THE NETWORK OPERATIONS CENTER OF THE NATIONAL RESEARCH COMPUTER NETWORK OF RUSSIA

A. G. Abramov and V. A. Porkhachev

St.-Petersburg Department of Joint Supercomputer Center of the Russian Academy of Sciences — Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences,” 14 Smolenka River Emb., St. Petersburg 195251, Russian Federation

Abstract: The paper is devoted to the general principles and methods of development and improvement of the software suite of the digital Network Operations Center of the new generation National Research Computer Network of Russia (NOC NIKS). The paper studies and systematizes the tasks solved by NOCs of large telecommunications networks including those specific to national research and education networks. Based on the results, the main tasks and functions of NOC NIKS are highlighted, the arguments and justifications for creating your own software solution are given, its novelty, distinctive competitive advantages, and unique functions are indicated, and possible directions for development and practical use are proposed. Special attention is paid to the presentation and discussion of the features of the architecture of the program core of the suite as well as interconnected basic and additional components including a set of functional and auxiliary modules as part of the core, microservices for interacting with external systems, and external services.

Keywords: National Research Computer Network, NIKS; Network Operations Center, NOC; network monitoring and management; software suite

DOI: 10.14357/08696527230304

EDN: PRLQXO

Acknowledgments

The paper has been prepared within the framework of the state assignment of FSI “Scientific Research Institute for System Analysis of the Russian Academy of Sciences” (topic No. FNEF-2022-0014).

References

1. Claise, B., and R. Wolter. 2012. *Network management: Accounting and performance strategies*. Cisco Press. 631 p.
2. Subramanian, M., T. A. Gonsalves, and N. Usha Rani. 2012. *Network management: Principles and practice*. 2nd ed. Prentice Hall. 695 p.
3. Ge, J., T. Li, and Y. Wu. 2022. *AI and machine learning for network and security management*. IEEE Press ser. on networks and service management. Wiley/IEEE Press. 304 p.

4. McClure, C. R., A. P. Bishop, P. Doty, and H. Rosenbaum. 1991. *The national research and education network (NREN): Research and policy perspectives*. Praeger. 758 p.
5. Allocchio, C., L. Balint, V. Berkhout, et al. 2010. *A history of international research networking: The people who made it happen*. New York, NY: Wiley/VCH. 317 p.
6. GÉANT: The case for NRENs. A repository of resources to support funding, advocacy and the advancement of national and regional R&E networks. Available at: <https://caseforrens.org> (accessed September 14, 2022).
7. Abramov, A. G., A. A. Gonchar, A. V. Evseev, and B. M. Shabanov. 2021. Natsional'naya issledovatel'skaya komp'yuternaya set' novogo pokoleniya: tekushchee sostoyanie i kontseptsiya razvitiya [The new generation National Research Computer Network: Current status and concept for the development]. *Informatsionnye tekhnologii* [Information Technologies] 27(3):115–124. doi: 10.17587/it.27.115-124.
8. Abramov, A. G., A. V. Evseev, A. A. Gonchar, and B. M. Shabanov. 2022. Voprosy uvelicheniya propusknoy sposobnosti i territorial'noy dostupnosti natsional'noy issledovatel'skoy komp'yuternoy seti Rossii [Issues of increasing the network bandwidth and territorial accessibility of National Research Computer Network of Russia]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 2(32):4–12. doi: 10.14357/08696527220201.
9. SIG-NOC Tools Survey: What software tools R&E operations centres use. Available at: <https://wiki.geant.org/display/SIGNOC/SIG-NOC+Tools+Survey+2019> (accessed September 14, 2022).
10. Abramov, A. G. 2021. Issues of modernization of the monitoring and control system of the National Research Computer Network of Russia with an emphasis on free software solutions. *Lobachevskii J. Mathematics* 42(11):2469–2480. doi: 10.1134/S1995080221110020.
11. Abramov, A. G., and V. A. Porkhachev. 2022. Architecture and key components of the improved digital center for monitoring and management of the National Research Computer Network of Russia. *Lobachevskii J. Mathematics* 43(10):2822–2832. doi: 10.1134/S199508022130029.

Received July 21, 2023

Contributors

Abramov Alexey G. (b. 1976) — Candidate of Science (PhD) in physics and mathematics, assistant professor, leading scientist, St.-Petersburg Department of Joint Supercomputer Center of the Russian Academy of Sciences — Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences,” 14 Smolenka River Emb., St. Petersburg 195251, Russian Federation; abramov@niks.su

Porkhachev Vasilii A. (b. 1967) — leading specialist, St.-Petersburg Department of Joint Supercomputer Center of the Russian Academy of Sciences — Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences,” 14 Smolenka River Emb., St. Petersburg 195251, Russian Federation; porhachev@niks.su

АНАЛИЗ ЭФФЕКТИВНОСТИ АЛГОРИТМА РЕДУКЦИИ В РЕШЕНИИ ЗАДАЧИ ОБ УПАКОВКЕ В КОНТЕЙНЕРЫ*

Е. Б. Барашов¹, А. В. Егоркин², Д. В. Лемтюжникова³, М. А. Посыпкин⁴

Аннотация: Задача упаковки в контейнеры — это известная комбинаторная задача, заключающаяся в поиске минимального числа контейнеров фиксированного размера для размещения заданного набора предметов с известным весом. Несмотря на простую формулировку, задача относится к NP-трудным, и точные методы ее решения зачастую неэффективны на практике. Поэтому большое значение имеет исследование и разработка приближенных методов решения задачи об упаковке. В статье исследуется класс приближенных алгоритмов, состоящих в последовательном применении редукции и одного из четырех «жадных» алгоритмов. Проводится оценка влияния редукций на качество получаемых решений и время работы рассматриваемых методов. Алгоритмы сравниваются на четырех наборах данных по нескольким критериям, отвечающим за качество получаемых решений и время, необходимое для их нахождения. Проведенное экспериментальное исследование показало, что эффективность применения редукции варьируется в широких пределах и сильно зависит от коэффициентов задачи.

Ключевые слова: редукция; задача упаковки в контейнеры; дискретная оптимизация; жадные алгоритмы; улучшение решения двусторонним помещением

DOI: 10.14357/08696527230305

EDN: QDUAMY

1 Введение

Задача упаковки в контейнеры (bin-packing problem, ВВР) — это классическая задача дискретной оптимизации. Она заключается в упаковке предметов разного размера в ограниченное число контейнеров фиксированной вместимости с целью минимизировать число используемых контейнеров.

Классическая постановка задачи об упаковке в контейнеры впервые сформулирована в работе [1], где помимо постановки приводятся примеры применения этой задачи для оптимального размещения данных в памяти программы. Одним из популярных обобщений классической постановки считается

* Исследование частично поддерживается Российским научным фондом (проект 22-71-10131).

¹ Институт проблем управления Российской академии наук, barashov.eb@gmail.com

² Московский институт электронной техники, egorkinandrewv@mail.ru

³ Институт проблем управления Российской академии наук; Московский авиационный институт, darabbt@gmail.com

⁴ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mrosyrkin@frccsc.ru

задача о геометрической упаковке, состоящая в минимизации числа квадратных контейнеров, нужных для упаковки прямоугольных деталей [2]. Данная задача является частным случаем более общей постановки многомерной (векторной) упаковки, где предметы и контейнеры имеют $d \geq 2$ измерений. Подробный обзор вариантов многомерной задачи об упаковке приведен в работе [3].

Задача упаковки в контейнеры имеет множество и других практических приложений. Применение методов решения задачи упаковки в контейнеры для построения оптимального расписания многопроцессорной системы рассматривается в работе [4]. В работе [5] ВРР применяется для решения задач оптимального расписания работы многопроцессорной системы и построения наилучшего распределения пассажиров по автобусам. Следует также отметить приложения задачи об оптимальной упаковке в области планирования хирургических операций [6] и облачных вычислений [7].

Задача ВРР относится к NP-трудным [8], и нахождение точного решения зачастую невозможно для постановок, возникающих на практике. Поэтому при практическом применении для решения задачи упаковки в контейнеры часто используются эвристические алгоритмы из-за их способности быстро находить приближенные решения, близкие к оптимальным. Одним из основополагающих и наиболее распространенных видов эвристических алгоритмов считаются так называемые *жадные алгоритмы*, основанные на принятии локально-оптимальных решений на каждом шаге. Базовые жадные эвристики NFD (next-fit decreasing), FFD (first-fit decreasing), BFD (best-fit decreasing) и WFD (worst-fit decreasing) известны давно (см., например, [9]).

Для сокращения перебора в комбинаторных задачах применяют *редукции* — правила, позволяющие определить значения некоторых переменных задачи, опираясь на ее свойства. Редукции хорошо зарекомендовали себя в комбинации с точными алгоритмами решения, например с методом ветвей и границ [9]. Как будет показано далее в статье, редукции, предварительно применяемые к исходной задаче, могут увеличивать эффективность жадных алгоритмов.

С целью определения наилучшего варианта сочетания редукции с различными эвристиками в статье проведено экспериментальное сравнение жадных эвристик и комбинации «редукция + эвристика» на нескольких тестовых наборах данных по нескольким критериям: качество работы редукции, число полученных оптимальных решений, среднее отклонение полученных приближенных решений от оптимумов (в процентах), среднее время работы. Результаты экспериментов показали, что эффект от предварительного применения редукции сильно зависит от исходных данных задачи (числа предметов в наборе, вместимости контейнера, диапазона весов предметов в наборе) и используемой эвристики.

Работа организована следующим образом. Второй раздел содержит обзор литературы, нацеленный на исследование применения редукции для ускорения работы алгоритмов решения задач дискретной оптимизации. Третий раздел содержит постановку задачи и описание используемых алгоритмов. В четвертом

разделе подробно описаны данные, использованные для экспериментов, полученные результаты и краткие выводы по каждому эксперименту. Наконец, в пятом разделе подводятся итоги экспериментов и предлагаются дальнейшее перспективы исследования.

2 Обзор литературы

Изучение процедуры редукции для ВРР можно проследить от основополагающей работы [8]. Процедуры редукции [10, 11] сыграли ключевую роль в разработке приближенных алгоритмов для ВРР. Основная идея процедуры редукции заключается в преобразовании примера задачи в более простой пример с сохранением оптимального значения решения. Затем этот более простой пример может быть решен более эффективно, обеспечивая полезную отправную точку для разработки приближенных алгоритмов.

На протяжении многих лет исследователи предлагали различные процедуры сокращения для ВРР, включая известный алгоритм FFD, который сортирует предметы в порядке уменьшения размера и затем упаковывает их жадным алгоритмом, используя первый доступный контейнер. Хотя FFD не оптимален, на практике он показал хорошую производительность и стал начальным алгоритмом для сравнения производительности других приближенных алгоритмов.

В статье [12] используется алгоритм FFD в качестве базового алгоритма в гиперэвристическом подходе, который объединяет несколько алгоритмов для решения задачи упаковки контейнера с целью улучшения результатов. В [13] авторы исследовали энергопотребление центра обработки данных с использованием CloudSim до и после применения алгоритмов упаковки в контейнеры: FFD и Ant Colony Optimization (ACO). Их результаты показали, что FFD был более эффективен, чем ACO, в минимизации энергопотребления для любого типа облачных систем. В [14] авторы используют два хорошо известных алгоритма для оптимизации энергопотребления систем реального времени, работающих на многоядерных платформах: BFD и WFD используются для сортировки задач в порядке убывания их размеров, а затем распределения их по доступным ядрам таким образом, чтобы свести к минимуму фрагментацию ядер. Авторы сравнивают предложенный подход с двумя другими методами разделения задач: фиксированным разделением и динамическим разделением. Результаты показывают, что предложенный подход обеспечивает более высокую энергоэффективность при одновременном удовлетворении ограничениям задачи.

Одна из основных проблем при разработке приближенных алгоритмов для задачи упаковки в контейнеры — получение жестких нижних границ для оптимального значения решения. Нижние границы важны, поскольку они обеспечивают эталон, по которому можно оценить эффективность приближенных алгоритмов. В работе [15] предложены новые нижние границы и процедуры

редукции для задачи упаковки в контейнеры, которые широко используются и изучаются с момента публикации статьи. Нижние границы основаны на идее разбиения предметов на подмножества и ограничении числа контейнеров, необходимых для упаковки каждого подмножества. Процедуры редукции предназначены для преобразования примера задачи в меньший, сохраняя при этом ее оптимальное решение. В работе [16] исследуются нижние оценки для ВВР, которые были введены в [15]. Показана взаимосвязь этих оценок. Аналитически получено значение наихудшего асимптотического коэффициента одной из оценок и показано, что оно неуплучшаемо.

В работе [17] вводятся новые классы быстрых нижних оценок для ВВР с улучшенной асимптотической производительностью в наихудшем случае. На большом наборе данных показано, что предложенные оценки уменьшают отклонение от оптимума и вычислительные затраты на решение. В работе [18] представлен общий подход к получению новых быстрых нижних оценок, основанный на двойственных допустимых функциях. Анализ наихудшего случая и результаты вычислений показывают, что один из предложенных классов превосходит наилучшую нижнюю оценку из [15]. Получено значение асимптотической производительности в худшем случае для оценки, которая может быть вычислена за линейное время для элементов, отсортированных по размеру.

Отдельный интерес представляют алгоритмы редукций для ВВР. В статье [19] предлагаются новые эвристики, основанные на концепции контроля среднего веса, который служит мерой среднего веса предметов в контейнерах. Авторы используют технику редукции для упрощения примера задачи и повышения эффективности эвристики. Техника редукции основана на том наблюдении, что некоторые предметы могут быть упакованы вместе без ущерба для оптимального решения. Авторы показывают, как определить такие группы предметов и упаковать их в один контейнер, что уменьшает размер задачи и делает эвристику более эффективной. Техника уменьшения применяется итеративно, и авторы показывают, что она приводит к значительному уменьшению размера задачи без ущерба для оптимального решения. Эвристики оцениваются на большом наборе эталонных примеров, и результаты показывают, что они превосходят современные эвристики как по качеству решения, так и по времени вычислений.

В статье [20] предлагается новая эволюционная эвристика для решения задачи упаковки в контейнеры. Эвристика основана на генетическом алгоритме, который использует набор эвристик для генерации новых решений и улучшения пригодности популяции. Для повышения эффективности генетического алгоритма авторы используют технику редукции, которая упрощает пример задачи путем выявления и удаления лишних элементов. Авторы показывают, что техника сокращения приводит к значительному уменьшению размера задачи, что повышает эффективность генетического алгоритма. Они также показывают, что используемая специальная редукция не влияет на качество решений, получаемых алгоритмом.

3 Базовая теория и алгоритмы

3.1 Постановка задачи

Приведем постановку одномерной задачи упаковки в контейнеры (BPP) [9]. Набор предметов w_1, \dots, w_n различных целочисленных размеров должен быть упакован в конечное число контейнеров (бинов), у каждого из которых ограниченная целочисленная вместимость c , таким образом, чтобы число использованных контейнеров было минимально. Предполагается, что $w_i \leq c$ для $i = 1, \dots, n$.

Несмотря на то что были разработаны методы точного решения BPP, на практике решающую роль играют эвристики, поскольку они способны справиться со сложными примерами за разумное время. В оставшейся части этого раздела приводится описание нескольких эвристик, предложенных в литературе. При описании будет использоваться термин «бин открывается» для ситуации, когда задействуется новый бин из числа неиспользованных, и «бин закрывается», когда бин далее не рассматривается для добавления в него новых предметов.

3.2 Жадные эвристики

Опишем известные классические жадные алгоритмы, приведенные в [9, 21]. Перед началом работы эвристик выполняется предварительная сортировка предметов в порядке убывания весов, поэтому в названиях всех эвристик присутствует буква 'D' — decreasing. Все жадные эвристики начинаются с того, что первый предмет помещается в первый бин.

NFD — next-fit-decreasing. Если в текущем бине достаточно свободного места, кладем предмет в него. В противном случае закрываем текущий бин, открываем новый, делаем его текущим и кладем в него предмет.

FFD — first-fit-decreasing. Каждый предмет помещается в первый бин, начиная с начала списка задействованных бинов, в котором достаточно места для его размещения. Если ни в одном из бинов места нет, открываем новый бин и кладем в него предмет.

BFD — best-fit-decreasing. Бин для размещения очередного предмета выбирается таким образом, чтобы оставшееся в бине пространство было минимальным. Если ни в одном из уже задействованных бинов места нет, открываем новый бин и помещаем предмет в него.

WFD — worst-fit-decreasing. Бин для размещения очередного предмета выбирается таким образом, чтобы оставшееся в бине пространство было максимальным. Если ни в одном из уже задействованных бинов места нет, открываем новый бин и кладем в него предмет.

TSFD — two-sides-fit-decreasing. Для каждого предмета ищется два набора элементов, которые могут быть помещены вместе с ним в бин. Первый набор составляется из предметов, начиная с начала списка (максимальные

по весу предметы), второй — с конца (содержит минимальные элементы). В бин помещается выбранный предмет и набор из двух рассмотренных, с наибольшим весом.

3.3 Алгоритмы редукции

Рассматриваемым алгоритмом редукции был выбран алгоритм, представленный в [9]. Он считается классическим и наиболее часто упоминается в литературе.

1. Выбрать предмет, который еще не был рассмотрен (предмет A).
2. Посчитать максимальное число предметов из оставшихся, которые можно положить вместе с выбранным (число k).
3. Если $k = 0$, предмет A помещается в бин, бин закрывается.
4. Если $k = 1$, выбирается максимальный предмет из оставшихся, который можно положить вместе с предметом A , помещают их вместе в бин, бин закрывается.
5. Если $k = 2$:
 - (а) найти предметы B и C , которые соответствуют следующим условиям:
 - (i) $w_A + w_B + w_C \leq c$;
 - (ii) $w_B + w_C$ максимальна;
 - iii) $C > B$;
 - (б) найти элемент D , для которого:
 - (i) $w_A + w_D \leq c$;
 - (ii) w_D максимален;
 - (в) если $w_D > w_B + w_C$, в бин помещаются предметы A и D и бин закрывается;
 - (г) в противном случае, если $w_D = w_B$ и выполнено хотя бы одно из соотношений
 - (i) $C - B \leq 2$;
 - (ii) $w_A + w_{C-1} + w_{C-2} > C$,в бин помещаются предметы A , B и C и бин закрывается.
6. Если еще не все предметы были рассмотрены, выполняется переход к шагу 1.

Известно, что применение редукции позволяет существенно сократить перебор в методе ветвей и границ [9]. В данной работе предлагается применить редукции для ускорения жадных алгоритмов. Для этого сначала к исходной задаче применяется редукция, а потом к редуцированной задаче — жадный алгоритм. Как будет показано в следующем разделе, такая комбинация приводит к существенному повышению качества работы алгоритма на ряде примеров.

4 Численные эксперименты

4.1 Среда и исходные данные для проведения экспериментов

Использован язык программирования Python (версия 3.9), библиотека `numpy` для работы с данными, библиотека `os` для операций ввода-вывода и библиотека `matplotlib` для построения графиков.

Для проведения экспериментов были использованы следующие наборы данных:

- `bin1data` — 36 различных классов, в каждом по 20 примеров¹;
- `bin2data` — 48 различных классов, в каждом по 10 примеров²;
- `bin3data` — 10 сложных примеров³;
- `hard28` — 28 очень сложных примеров⁴.

Набор `hard28` считается классическим набором для тестирования алгоритмов, применяемых для решения задачи ВРР, что отражено во многих статьях. Для более комплексного анализа были также взяты три больших набора (`bin1data`, `bin2data` и `bin3data`), представленных на сайте Йенского университета. Для этих наборов характерно большое число различных вариантов условий задачи с различными значениями вместимости корзин, количества и распределения весов предметов.

Параметры наборов данных приведены в табл. 1.

Таблица 1 Описание наборов данных `bin1data` и `bin2data`

Набор данных	Число предметов	Вместимость бина	Весы предметов
<code>bin1data</code>	50, 100, 200, 500	100, 120, 150	[1, 100], [20, 100], [30, 100]
<code>bin2data</code>	50, 100, 200, 500	1000	[13, 627]*

*Средние веса предметов — $c/3$, $c/5$, $c/7$ и $c/9$, где c — вместимость бина. Максимальное отклонение веса предмета от среднего веса — 20%, 50% и 90%.

4.2 Результаты экспериментов

Эксперименты показали, что редукция для одних данных улучшает работу жадных алгоритмов, а для других, напротив, ухудшает. Среди решенных примеров есть такие, для которых только с помощью одной лишь редукции можно решить задачу полностью (табл. 2). На наборе `bin1data` редукция позволяет сократить размерность на всех примерах, полностью решая часть из них. На

¹<https://www.wiwi.uni-jena.de/Entscheidung/binpp/bin1dat.htm>

²<https://www.wiwi.uni-jena.de/Entscheidung/binpp/bin2dat.htm>

³<https://www.wiwi.uni-jena.de/Entscheidung/binpp/bin3dat.htm>

⁴<https://www.euro-online.org/websites/esicup/wp-content/uploads/sites/12/2017/04/hard28.zip>

Таблица 2 Описание наборов данных `bin3data` и `hard28`

Набор данных	Число предметов	Вместимость бина	Веса предметов
<code>bin3data</code>	200	100 000	[20 000, 35 000]
<code>hard28</code>	[140, 200]	1000	[1,800]

Таблица 3 Эффективность применения редукции

Набор данных	Редуцированные предметы		
	Все предметы	Часть предметов	Ни один предмет
<code>bin1data</code>	258	462	0
<code>bin2data</code>	0	36	444
<code>bin3data</code>	0	0	10
<code>hard28</code>	0	28	0

Таблица 4 Доля примеров, в которых получено оптимальное решение

Набор данных	С редукцией	Без редукции
<code>bin1data</code>	562	488
<code>bin2data</code>	237	237
<code>bin3data</code>	0	0
<code>hard28</code>	5	5

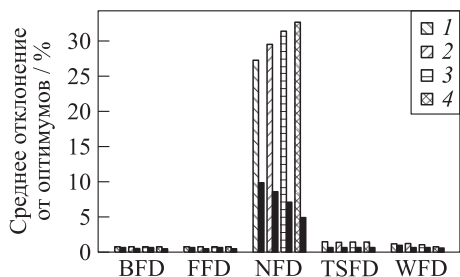
наборе `bin2data` только в 36 из 480 редукция приводит к уменьшению числа переменных. На наборе `bin3data` редукция не сработала ни на одном примере, а на `hard28` определенное уменьшение числа переменных произошло для каждого примера (табл. 3).

Также исследовалось влияние редукции на число примеров, в которых в результате применения эвристик были получены оптимальные решения (табл. 4). На наборе `bin1data` было получено больше оптимумов при работе редукции, на всех остальных — то же самое количество.

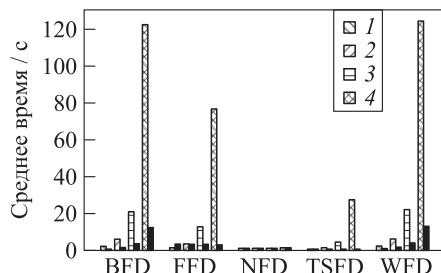
Важной характеристикой эффективности алгоритма служит близость получаемых приближенных решений к оптимальным. На рис. 1 показаны относительные отклонения полученных приближенных решений от оптимальных при запуске эвристик в комбинации с редукцией и без нее. Для вычисления относительного отклонения Δ использовалась формула:

$$\Delta = 100 \cdot \frac{f(x^r) - f(x^*)}{f(x^*)},$$

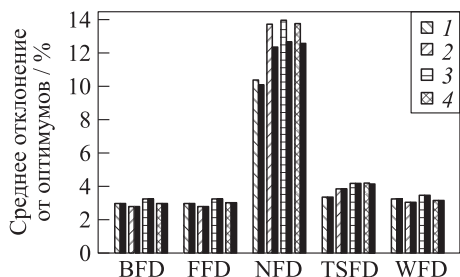
где x^* — оптимальное решение; x^r — решение, найденное эвристикой.



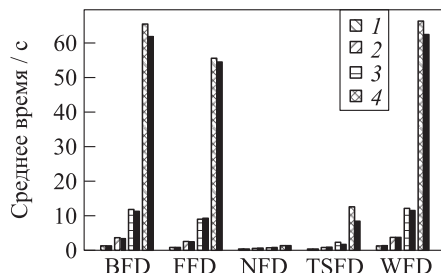
(a)



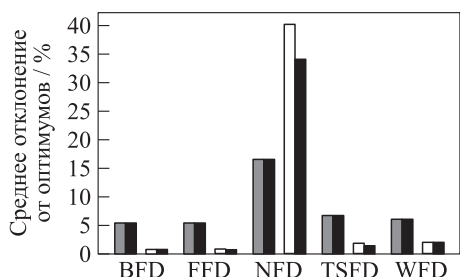
(a)



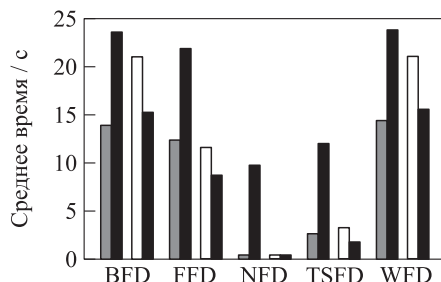
(б)



(б)



(в)



(в)

Рис. 1 Среднее отклонение от оптимумов полученных рекордов на редуцированных (черная заливка) и нередуцированных данных для различных количествах предметов: 1 — 50; 2 — 100; 3 — 200; 4 — 500; (a) bin1data; (б) bin2data; (в) bin3data (серая заливка); и hard28 (без заливки)

Рис. 2 Среднее время работы эвристики с применением редукции (черная заливка) и без применения для различных количествах предметов: 1 — 50; 2 — 100; 3 — 200; 4 — 500; (a) bin1data; (б) bin2data; (в) bin3data (серая заливка) и hard28 (без заливки)

На рис. 2 показано общее время работы программы с применением и без применения алгоритма редукции. Ситуация получилась неоднозначная: в одних случаях имеет место уменьшение времени работы при редуцировании, в других — увеличение. Например, на наборе данных bin3data, где редукция была

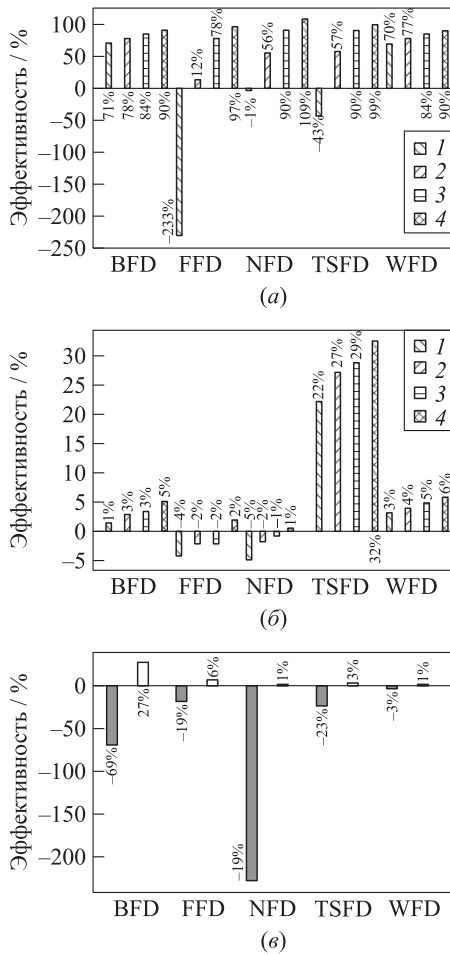


Рис. 3 Эффективность работы редукции для различных количеств предметов: 1 — 50; 2 — 100; 3 — 200; 4 — 500; (а) bin1data; (б) bin2data; (в) bin3data (серая заливка) и hard28 (без заливки)

и bin2data, но на более сложных задачах из наборов bin3data и hard28 оказалась менее эффективной. При этом при росте числа переменных эффект от применения редукции возрастает. Так как при применении редукции получаемые решения не ухудшаются, отрицательная эффективность показывает проигрыш по времени. Этот эффект хорошо заметен на наборе bin3data, где редукция не сработала, но было потрачено время на ее работу.

применена, она не сработала и получился значительный проигрыш во времени на каждом наборе для всех эвристик.

Для того чтобы соотнести выигрыш в точности решения к дополнительным временным затратам на его получение, была введена метрика эффективности применения редукции μ :

$$\mu = \Delta_f - \Delta_t.$$

При этом величина Δ_f , показывающая прирост точности за счет применения редукции, определяется следующим образом:

$$\Delta_f = 100\% \cdot \frac{f(x^G) - f(x^R)}{f(x^G)},$$

где x^R — решение, полученное с использованием совмещения редукции и жадного алгоритма; x^G — решение, полученное жадным алгоритмом без предварительной редукции. Величина Δ_t , отражающая увеличение времени, задается формулой:

$$\Delta_t = 100\% \cdot \frac{t_R - t}{t},$$

где t_R — время работы варианта, в котором совмещены редукция и жадный алгоритм; t — время работы жадного алгоритма, применяемого без редукции.

На рис. 3 изображено влияние редукции. Она позволяет улучшить результаты на наборах bin1data

5 Заключение

В данной работе исследовалось влияние редукции на время работы жадных алгоритмов, а также на точность полученных решений для задачи упаковки в контейнеры. Были реализованы несколько известных жадных алгоритмов и предложенная авторами эвристика TSFD. Все алгоритмы экспериментально сравнивались на большом наборе бенчмарков по следующим критериям: качество работы редукции; число полученных оптимальных решений; среднее отклонение от оптимумов (в процентах); среднее время работы алгоритма на редуцированных и нередуцированных данных.

При применении редукции вместе с быстрой, но неточной эвристикой можно получить более качественные результаты, однако время на подсчет может увеличиться. При применении же редукции с более точной эвристикой качество результата либо улучшится, либо останется таким же. Однако время подсчета может как увеличиться, так и уменьшиться. Если на наборе данных алгоритм редукции не сработал, наблюдается только увеличение общего времени работы программы. Эффективность применения алгоритма редукции сильно растет с увеличением числа предметов в наборе, так как при большом числе предметов редукция имеет больший шанс сработать и уменьшить задачу до запуска эвристики. Также если набор имеет большое число предметов с весом порядка 70% от вместимости корзины и более, то редукция может сильно уменьшить задачу, поскольку алгоритм ее работы позволяет эффективно работать с подобными задачами.

Эксперименты показали, что производительность алгоритмов зависит от коэффициентов задачи, а редукция не для всех данных улучшает работу алгоритмов. Комбинация эвристики и редукции, которая хорошо работает для одного примера задачи, может показать плохие результаты для другого примера. Таким образом, важно выявить эту зависимость и разработать некоторые правила, помогающие выбрать соответствующий алгоритм на основе коэффициентов задачи. Важным направлением представляется увеличение числа исследуемых эвристических алгоритмов и тестовых примеров, с тем чтобы исследовать на них влияние на качество работы алгоритмов других вариантов редукций и сделать сравнение более полным. В дальнейшем будут рассмотрены другие алгоритмы редукции для выявления наиболее качественного.

Литература

1. Garey M. R., Graham R. L., Ullman J. D. Worst-case analysis of memory allocation algorithms // 4th Annual ACM Symposium on Theory of Computing Proceedings. — New York, NY, USA: Association for Computing Machinery, 1972. P. 143–150. doi: 10.1145/800152.804907.
2. Ram B. The pallet loading problem: A survey // Int. J. Prod. Econ., 1992. Vol. 28. No. 2. P. 217–225. doi: 10.1016/0925-5273(92)90034-5.

3. *Christensen H. I., Khan A., Pokutta S., Tetali P.* Multidimensional bin packing and other related problems: A survey // *Computer Science Review*, 2017. Vol. 24. P. 63–79.
4. *Coffman Jr. E. G., Garey M. R., Johnson D. S.* An application of bin-packing to multiprocessor scheduling // *SIAM J. Comput.*, 1978. Vol. 7. No. 1. P. 1–17. doi: 10.1137/0207001.
5. *Ojeyinka T. O.* Bin packing algorithms with applications to passenger bus loading and multiprocessor scheduling problems // *Communications Applied Electronics*, 2015. Vol. 2. No. 8. P. 38–44. doi: 10.5120/cae2015651851.
6. *Vijayakumar B., Parikh P. J., Scott R., Barnes A., Gallimore J.* A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital // *Eur. J. Oper. Res.*, 2013. Vol. 224. No. 3. P. 583–591. doi: 10.1016/j.ejor.2012.09.010.
7. *Ye D., Xie F., Zhang G.* Truthful mechanism design for bin packing with applications on cloud computing // *J. Comb. Optim.*, 2022. Vol. 44. P. 2224–2245. doi: 10.1007/s10878-020-00601-4.
8. *Garey M. R., Johnson D. S.* Computers and intractability. A guide to the theory of NP-completeness. — New York, NY, USA: W. H. Freeman & Co., 1979. 340 p.
9. *Martello S., Toth P.* Knapsack problems: Algorithms and computer implementations. — John Wiley & Sons, Inc., 1990. 296 p.
10. *Alvim A. C. F., Ribeiro C. C., Glover F., Aloise D. J.* A hybrid improvement heuristic for the one-dimensional bin packing problem // *J. Heuristics*, 2004. Vol. 10. P. 205–229. doi: 10.1023/B:HEUR.0000026267.44673.ed.
11. *Carlier J., Clautiaux F., Moukrim A.* New reduction procedures and lower bounds for the two-dimensional bin packing problem with fixed orientation // *Comput. Oper. Res.*, 2007. Vol. 34. No. 8. P. 2223–2250. doi: 10.1016/j.cor.2005.08.012.
12. *Lopez-Camacho E., Terashima-Marin H., Ross P., Ochoa G.* A unified hyper-heuristic framework for solving bin packing problems // *Expert Syst. Appl.*, 2014. Vol. 41. No. 15. P. 6876–6889. doi: 10.1016/j.eswa.2014.04.043.
13. *Bouaouda A., Afdel K., Abounacer R.* Forecasting the energy consumption of cloud data centers based on container placement with ant colony optimization and bin packing // 5th Conference on Cloud and Internet of Things. — Piscataway, NJ, USA: IEEE, 2022. P. 150–157. doi: 10.1109/CIoT53061.2022.9766522.
14. *El Sayed M. A., Saad E. S. M., Aly R. F., Habash S. M.* Energy-efficient task partitioning for real-time scheduling on multi-core platforms // *Computers*, 2021. Vol. 10. No. 1. Art. 10. doi: 10.3390/computers10010010.
15. *Martello S., Toth P.* Lower bounds and reduction procedures for the bin packing problem // *Discrete Appl. Math.*, 1990. Vol. 28. No. 1. P. 59–70. doi: 10.1016/0166-218X(90)90094-S.
16. *Bourjolly J. M., Rebetz V.* An analysis of lower bound procedures for the bin packing problem // *Comput. Oper. Res.*, 2005. Vol. 32. No. 3. P. 395–405. doi: 10.1016/S0305-0548(03)00244-2.
17. *Crainic T. G., Perboli G., Pezzuto M., Tadei R.* New bin packing fast lower bounds // *Comput. Oper. Res.*, 2007. Vol. 34. No. 11. P. 3439–3457. doi: 10.1016/j.cor.2006.02.007.
18. *Fekete S. P., Schepers J.* New classes of fast lower bounds for bin packing problems // *Math. Program.*, 2001. Vol. 91. P. 11–31. doi: 10.1007/s101070100243.

19. *Fleszar K., Charalambous C.* Average-weight-controlled bin-oriented heuristics for the one-dimensional bin-packing problem // *Eur. J. Oper. Res.*, 2011. Vol. 210. No. 2. P. 176–184. doi: 10.1016/j.ejor.2010.11.004.
20. *Stawowy A.* Evolutionary based heuristic for bin packing problem // *Comput. Ind. Eng.*, 2008. Vol. 55. No. 2. P. 465–474. doi: 10.1016/j.cie.2008.01.007.
21. *Coffman E. G., Martello S., Csirik J., Vigo D., Galambos G.* Bin packing approximation algorithms: Survey and classification // *Handbook of combinatorial optimization*. — New York, NY, USA: Springer Science + Business Media, 2013. P. 455–531.

Поступила в редакцию 02.02.23

EFFICIENCY OF THE REDUCTION ALGORITHMS IN THE BIN PACKING PROBLEM

*E. B. Barashov*¹, *A. V. Egorkin*², *D. V. Lemtyuzhnikova*^{1,3}, and *M. A. Posypkin*⁴

¹V. A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation

²National Research University of Electronic Technology (MIET), 1 Shokina Sq., Moscow, Zelenograd 124498, Russian Federation

³Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

⁴Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The bin packing problem is a well-known combinatorial problem consisting in finding the minimum number of fixed-size bins to hold a given set of items with known weight. Despite its simple formulation, the problem is NP-hard and exact methods for its solution are often inefficient in practice. Therefore, the research and development of approximate methods for solving the bin packing problem is of great importance. The paper investigates a class of approximate algorithms consisting of the sequential application of reduction and one of four “greedy” algorithms. The effect of reduction on the quality of the solutions obtained and the running time of the methods under consideration is evaluated. The algorithms are compared on four data sets according to several criteria responsible for the quality of the obtained solutions and the time required to find them. The experimental study shows that the efficiency of the reduction varies widely and depends strongly on the problem coefficients.

Keywords: reduction; bin-packing problem; discrete optimization; greedy algorithms; solution improvement by two-sided placement

DOI: 10.14357/08696527230305

EDN: QDUAMY

Acknowledgments

This research is partially supported by the Russian Science Foundation (project No. 22-71-10131).

References

1. Garey, M. R., R. L. Graham, and J. D. Ullman. 1972. Worst-case analysis of memory allocation algorithms. *4th Annual ACM Symposium on Theory of Computing Proceedings*. New York, NY: Association for Computing Machinery. 143–150. doi: 10.1145/800152.804907.
2. Ram, B. 1992. The pallet loading problem: A survey. *Int. J. Prod. Econ.* 28(2):217–225. doi: 10.1016/0925-5273(92)90034-5.
3. Christensen, H. I., A. Khan, S. Pokutta, and P. Tetali. 2017. Multidimensional bin packing and other related problems: A survey. *Computer Science Review* 24:63–79.
4. Coffman, Jr. E. G., M. R. Garey, and D. S. Johnson. 1978. An application of bin-packing to multiprocessor scheduling. *SIAM J. Comput.* 7(1):1–17. doi: 10.1137/0207001.
5. Ojeyinka, T. O. 2015. Bin packing algorithms with applications to passenger bus loading and multiprocessor scheduling problems. *Communications Applied Electronics* 2(8):38–44. doi: 10.5120/cae2015651851.
6. Vijayakumar, B., P. J. Parikh, R. Scott, A. Barnes, and J. Gallimore. 2013. A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *Eur. J. Oper. Res.* 224(3):583–591. doi: 10.1016/j.ejor.2012.09.010.
7. Ye, D., F. Xie, and G. Zhang. 2022. Truthful mechanism design for bin packing with applications on cloud computing. *J. Comb. Optim.* 44:2224–2245. doi: 10.1007/s10878-020-00601-4.
8. Garey, M. R., and D. S. Johnson. 1979. *Computers and intractability. A guide to the theory of NP-completeness*. New York, NY: W. H. Freeman & Co. 340 p.
9. Martello, S., and P. Toth. 1990. *Knapsack problems: Algorithms and computer implementations*. John Wiley & Sons, Inc. 296 p.
10. Alvim, A. C. F., C. C. Ribeiro, F. Glover, and D. J. Aloise. 2004. A hybrid improvement heuristic for the one-dimensional bin packing problem. *J. Heuristics* 10:205–229. doi: 10.1023/B:HEUR.0000026267.44673.ed.
11. Carlier, J., F. Clautiaux, and A. Moukrim. 2007. New reduction procedures and lower bounds for the two-dimensional bin packing problem with fixed orientation. *Comput. Oper. Res.* 34(8):2223–2250. doi: 10.1016/j.cor.2005.08.012.
12. Lopez-Camacho, E., H. Terashima-Marin, P. Ross, and G. Ochoa. 2014. A unified hyper-heuristic framework for solving bin packing problems. *Expert Syst. Appl.* 41(15):6876–6889. doi: 10.1016/j.eswa.2014.04.043.
13. Bouaouda, A., K. Afdel, and R. Abounacer. 2022. Forecasting the energy consumption of cloud data centers based on container placement with ant colony optimization and bin packing. *25th Conference on Cloud and Internet of Things*. Piscataway, NJ: IEEE. 150–157. doi: 10.1109/CIoT53061.2022.9766522.
14. El Sayed, M. A., E. S. M. Saad, R. F. Aly, and S. M. Habash. 2021. Energy-efficient task partitioning for real-time scheduling on multi-core platforms. *Computers* 10(1):10. doi: 10.3390/computers10010010.

15. Martello, S., and P. Toth. 1990. Lower bounds and reduction procedures for the bin packing problem. *Discrete Appl. Math.* 28(1):59–70. doi: 10.1016/0166-218X(90)90094-S.
16. Bourjolly, J. M., and V. Reboz. 2005. An analysis of lower bound procedures for the bin packing problem. *Comput. Oper. Res.* 32(3):395–405. doi: 10.1016/S0305-0548(03)00244-2.
17. Crainic, T. G., G. Perboli, M. Pezzuto, and R. Tadei. 2007. New bin packing fast lower bounds. *Comput. Oper. Res.* 34(11):3439–3457. doi: 10.1016/j.cor.2006.02.007.
18. Fekete, S. P., and J. Schepers. 2001. New classes of fast lower bounds for bin packing problems. *Math. Program.* 91:11–31. doi: 10.1007/s101070100243.
19. Fleszar, K., and C. Charalambous. 2011. Average-weight-controlled bin-oriented heuristics for the one-dimensional bin-packing problem. *Eur. J. Oper. Res.* 210(2):176–184. doi: 10.1016/j.ejor.2010.11.004.
20. Stawowy, A. 2008. Evolutionary based heuristic for bin packing problem. *Comput. Ind. Eng.* 55(2):465–474. doi: 10.1016/j.cie.2008.01.007.
21. Coffman, E. G., S. Martello, J. Csirik, D. Vigo, and G. Galambos. 2013. Bin packing approximation algorithms: Survey and classification. *Handbook of combinatorial optimization*. New York, NY: Springer Science + Business Media. 455–531. doi: 10.1007/978-1-4419-7997-1_35.

Received February 2, 2023

Contributors

Barashov Egor B. (b. 1999) — engineer, V. A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; barashov.eb@gmail.com

Egorkin Andrew V. (b. 2001) — bachelor student, National Research University of Electronic Technology (MIET), 1 Shokina Sq., Moscow, Zelenograd 124498, Russian Federation; egorkinandrewv@mail.ru

Lemtyuzhnikova Darya V. (b. 1988) — Candidate of Science (PhD) in physics and mathematics, associate professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; senior scientist, V. A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; darabbt@gmail.com

Posypkin Mikhail A. (b. 1974) — Doctor of Science in physics and mathematics, Corresponding Member of the Russian Academy of Sciences, deputy director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; mposypkin@frccsc.ru

МЕТОДОЛОГИЯ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ ПРИ СОЗДАНИИ ФУНКЦИЙ СИСТЕМ СКАНИРОВАНИЯ И ПЕЧАТИ

И. В. Сафонов¹, И. А. Матвеев²

Аннотация: Устройства сканирования, копирования и печати широко используются в промышленности, государственном управлении и для персональных нужд. В сочетании с прикладным и системным программным обеспечением (ПО) такие устройства представляют собой системы сканирования и печати. Формулировка методологии научно-исследовательской работы (НИР) при создании функций подобных систем позволяет формализовать и ускорить разработку новых методов обработки изображений. Приводится обзор методологий научного исследования, НИР в промышленности, разработки ПО. Для НИР при создании решений для систем сканирования печати в качестве критерия успеха выбран факт реализации результата исследовательской работы в продукте. На основе ретроспективного анализа выявлены факторы, которые препятствовали реализации результатов. На стадии НИР данные факторы можно рассматривать как риски. С помощью метода анализа иерархий риски ранжированы по важности. Предложена методология НИР при создании новых функций систем сканирования и печати, основанная на анализе и управлении рисками.

Ключевые слова: методология НИР; управление рисками; метод анализа иерархий

DOI: 10.14357/08696527230306

EDN: QEYMKА

1 Введение

Несмотря на существующую тенденцию к переходу на электронный документооборот, офисные устройства печати — принтеры — широко используются как в промышленности и делопроизводстве, так и для персональных нужд [1]. В настоящее время имеет место переход от принтеров, используемых исключительно для печати и реализованных как периферийные устройства персональных компьютеров (ПК), к многофункциональным устройствам (МФУ), или многофункциональным принтерам (МФП), позволяющим выполнять печать, сканирование и копирование. Многофункциональное устройство способно работать как при подключении к ПК, так и в качестве самостоятельного устройства. В сочетании с прикладным и системным ПО современные МФП представляют собой системы сканирования и печати.

¹Национальный исследовательский ядерный университет «МИФИ», ilia.safonov@gmail.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, matveev@frccsc.ru

Рост мощности вычислительных платформ и бурное развитие методов цифровой обработки изображений, компьютерного зрения и распознавания образов привели к появлению большого числа программ и программно-аппаратных комплексов для обработки фотографий и изображений различного рода документов. Однако их применение в системах печати часто затруднено в силу существенно меньших вычислительных ресурсов по сравнению с ПК и спецификой процесса обработки информации. Предпочтительной выглядит разработка таких новых функций МФП, которые могли бы быть реализованы без изменения аппаратной части за счет модификации системного и/или встроенного ПО.

Формулировка методологии НИР при создании решений для МФП позволяет формализовать и ускорить разработку методов обработки изображений, что в итоге ведет к внедрению новых функций информационных систем сканирования и печати.

2 Существующие методологии исследований и разработки

Методология — это учение о структуре, логической организации, методах и средствах деятельности [2]. Дадим краткий обзор методологий, которые позволяют организовать и формализовать деятельность по решению задач исследований и разработки. Начнем с перечисления этапов научного исследования, затем рассмотрим этапы НИР в промышленности, завершим методологиями разработки ПО. Выделяют следующие этапы научного исследования [3]:

- выбор темы и обоснование ее актуальности;
- ознакомление с проблемой посредством обзора существующей информации по данной теме;
- уточнение темы и формулировка плана НИР;
- собственно исследование, содержание которого зависит от используемых методов (экспериментальные, теоретические и т. д.);
- обработка и анализ результатов исследования, формулировка выводов;
- оформление документации с результатами исследования.

ГОСТ 15.101-2021 [4] определяет следующие общие этапы НИР при разработке продукции:

- выбор направления исследований на основе анализа современного состояния проблемы, в том числе результатов поиска объектов интеллектуальной собственности по данной теме;
- теоретические и экспериментальные исследования для решения поставленных задач;
- обобщение и оценка результатов исследований, выпуск научно-технической документации;

– приемка НИР в соответствии с критериями, определенными в техническом задании.

Для разработки ПО предложено множество методологий [5]. Классическая каскадная методология («водопад», waterfall) состоит из следующих этапов: анализ проекта; анализ требований; проектирование продукта; разработка продукта; системное тестирование; поддержка. Каждый последующий этап стартует только по окончании предыдущего. Предполагается, что все требования к итоговому продукту известны заранее, неопределенность отсутствует, никаких новых запросов на изменения в ходе проекта не произойдет. Такие предположения выглядят идеализированными, поэтому в настоящее время классическая каскадная методология применяется редко, зато возникло множество вариантов ее развития. Одним из таких вариантов стала итеративная каскадная методология, в которой возможен возврат на предыдущие этапы. Еще один вариант — это инкрементная каскадная методология, где требования по функциональности ПО возрастают от версии к версии, а каждая версия разрабатывается, например, по итеративной каскадной методологии. Каскадные методологии разработки призваны контролировать сроки и бюджет выполнения проектов по разработке ПО, но их успешное применение возможно при условии высокой определенности с требованиями на старте проекта, что для исследовательских задач нехарактерно.

Спиральная методология основана на анализе рисков [6]. Риски — это любые неопределенности и события, которые могут помешать достичь целей проекта. В данной методологии циклически повторяются следующие основные этапы: определение задач, альтернатив и ограничений; оценка альтернатив, выделение рисков и способов борьбы с ними; разработка и верификация очередной версии ПО; планирование следующей итерации. На каждом витке спирали создается прототип продукта. Успех проекта во многом определяется качеством анализа и управления рисками.

В настоящее время массово используются разновидности гибкой (agile) методологии разработки ПО [7], которые предусматривают постоянное привлечение заказчика или пользователей для оценки текущих результатов, формулировки задач и расстановки приоритетов. Планирование осуществляется на короткие промежутки времени, так называемые спринты. Стремятся в конце каждого спринта иметь работоспособный прототип или версию ПО.

Как правило, на практике та или иная методология разработки ПО используется не в классическом варианте, а в модифицированном виде с учетом особенностей решаемых задач и структуры организации, часто применяются комбинации элементов нескольких методологий.

3 Методология научно-исследовательской работы при создании решений для принтерных систем

Предлагаемая методология стала результатом индуктивного обобщения опыта работы исследовательских подразделений компаний-производителей электрони-

Риски НИР при создании решений для МФП

Обозначение	Риск	Вес
ρ_1	Функция не привлекательна для пользователей	0,190
ρ_2	Нарушается чужая интеллектуальная собственность	0,182
ρ_3	Невозможно реализовать на целевой платформе	0,141
ρ_4	Неудовлетворительное качество	0,139
ρ_5	Ухудшаются показатели производительности	0,138
ρ_6	Используется ПО с неподходящей лицензией	0,059
ρ_7	Высокие затраты на реализацию в продукте	0,045
ρ_8	Качество хуже, чем у конкурентов	0,045
ρ_9	Длительные сроки реализации в продукте	0,039
ρ_{10}	Вопрос защиты интеллектуальной собственности не решен	0,020

ки [8, 9]. Она сочетает в себе ряд этапов методологий научного исследования, НИР в промышленности и разработки ПО, которые дали положительный эффект при разработке методов обработки изображений для МФП.

Во многих компаниях процесс разработки новых функций для устройств и ПО построен следующим образом: сначала исследовательским подразделением выполняется НИР, затем результаты НИР всесторонне оцениваются и при условии положительных оценок передаются в производственные подразделения для реализации в продукте. Критерии успешности выполнения проектов НИР, как правило, определены достаточно размыто. Будем считать НИР успешной, если ее результаты были реализованы в коммерческом продукте, т. е. каком-либо МФП или поставляемом с ним ПО. Основываясь на данном критерии, нельзя сделать вывод об успешности проекта сразу по его окончании, необходимо время, чтобы понять, нашли результаты НИР практическое применение или нет.

Ретроспективный анализ около 40 проектов НИР позволил выявить факторы, которые воспрепятствовали реализации в продуктах их результатов. На стадии выполнения проекта эти факторы можно считать рисками, которые еще не имели места. Минимизация рисков повышает вероятность успешности НИР. Таким образом, целесообразно формулировать методологию НИР на основе анализа и управления рисками. В таблице перечислены риски НИР при создании решений для МФП. Риски ранжированы с помощью метода анализа иерархий (МАИ, analytic hierarchy process, АНР) [10], для каждого риска определен вес, характеризующий его относительную важность. По весу можно выделить три группы рисков: ρ_1 – ρ_2 ведут к отказу от реализации результатов НИР; ρ_3 – ρ_5 ведут либо к отказу от реализации, либо к тому, что принятие решения будет отложено на несколько лет; ρ_6 – ρ_{10} , вероятно, повлекут задержку с реализацией от нескольких месяцев до нескольких лет.

Предлагается основанная на анализе и управлении рисками методология НИР при создании решений для МФП. Базовая идея похожа на спиральную методологию разработки ПО, но содержание этапов отличается. Схема мето-

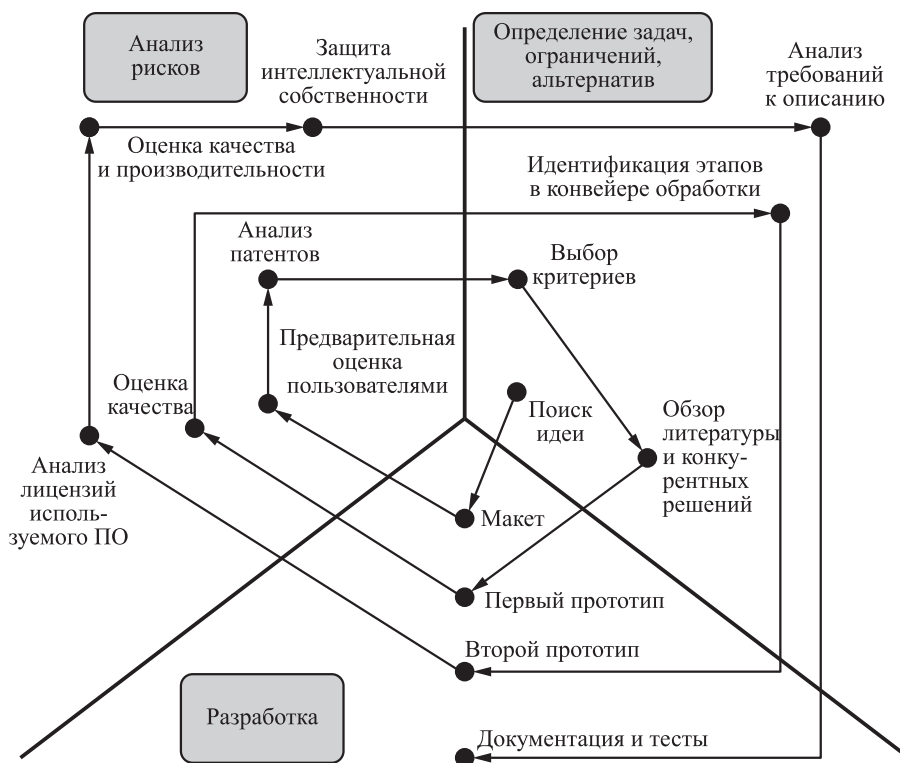


Схема методологии НИР при создании решений для МФП

логии приведена на рисунке. Подразумевается, что последовательность этапов делает три витка, переходя от определения задач, ограничений и альтернатив к анализу рисков, а затем к разработке. Если при оценке риска выясняется, что он достаточно высок, то возможно полное либо частичное повторение предыдущего витка при условии, что это не противоречит срокам выполнения НИР.

Проект начинается с поиска идеи новой функции для систем сканирования и печати. На данном этапе изучается продукция конкурентов, анализируются задачи, возникающие у пользователей, и рассматриваются возможные альтернативы. Далее создается чрезвычайно упрощенный макет продукта, чтобы подтвердить или опровергнуть гипотезу, что пользователи будут успешно применять продукт в данном виде. В работе [10] для подобного рода макетов вводится термин «прототип». В области проектирования визуальных интерфейсов пользователей аналогом «прототипа» служит макет с низкой детализацией (low-fidelity mockup), который позволяет оценить, насколько интерфейс будет понятен и удобен [11]. Однако концепция «прототипа» шире, так как позволяет оценить не только интерфейс пользователя, но и иные характеристики продукта,

например качество работы ПО оптического распознавания символов для текста, напечатанного в экономичном режиме [9]. Поскольку термин «прототип» не получил широкого распространения, будем использовать термин «макет». При разработке методов обработки изображений подобный макет — это, как правило, вручную нарисованное или обработанное изображение.

На следующем этапе для макета выполняется предварительная оценка пользовательских свойств для минимизации рисков ρ_1 и ρ_4 . Если макет пользователям понравился, то для управления риском ρ_2 выполняется переход к поиску и анализу патентной информации по данной тематике. Заключительным этапом первого витка спирали становится выбор критериев оценки качества разрабатываемого метода, подготовка тестовых данных и формулировка задачи исследования. Критерии должны быть количественными, причем предпочтительно использовать критерии, известные в индустрии печати или похожие на таковые.

Второй виток спирали начинается с обзора литературы и решений конкурентов по выбранной теме. Этот этап характерен для любого научного исследования. Его целью ставится поиск подходов к решению задачи НИР. Подход, который рассматривается как наиболее перспективный, применяется для разработки алгоритма и первого прототипа ПО. Как правило, первый прототип программы разрабатывается с помощью высокоуровневых средств, например MATLAB [12] или пакетов Python для обработки изображений, сигналов, машинного обучения и т. п. Использование высокоуровневых средств позволяет быстро получить прототип для оценки качества работы разработанного алгоритма.

На следующем этапе путем тестирования первого прототипа выполняется оценка качества работы метода для минимизации рисков ρ_4 и ρ_8 . Заметим, что время обработки на данном этапе не оценивается. Если качество разработанного метода удовлетворительное, то выполняется идентификация этапов в конвейере печати или сканирования, подходящих для реализации данного метода. Такой поиск альтернатив предназначен для нахождения запасных вариантов на случай, если реализация метода на целевой платформе окажется невозможна, или его производительность будет низкой, или стоимость и время реализации окажутся высокими. Данная деятельность, начинающая третий виток, направлена на управление рисками ρ_3 , ρ_5 , ρ_7 и ρ_9 .

Одним из результатов этапа идентификации возможных мест для реализации метода становятся требования ко второму прототипу ПО, который реализуется с помощью языков программирования C/C++ с учетом возможностей и ограничений целевых вычислительных платформ. Далее для минимизации риска ρ_6 выполняется анализ лицензий используемых библиотек. Затем оцениваются качество и время работы второго прототипа, что направлено на управление рисками ρ_3 , ρ_4 , ρ_5 и ρ_8 . Если сформулированные на первом витке критерии удовлетворены, то для управления риском ρ_{10} осуществляется переход к анализу подходов к защите разработанного метода как объекта интеллектуальной собственности. Завершают третий виток анализ требований производственных подразделений к документированию и оформлению кода, а также подготовка

документации и тестов в соответствии с этими требованиями. Эти этапы предназначены для минимизации рисков ρ_7 и ρ_9 .

4 Заключение

Предложена методология НИР по созданию новых функций систем сканирования и печати, основанная на анализе рисков и управлении ими. Эта методология представляет собой оригинальную разработку авторов, основанную на индуктивном обобщении опыта работы и использовании ряда элементов существующих методологий исследований и разработки. Предложенная методология НИР с небольшими модификациями может быть применена в промышленности при создании решений в сходных с рассматриваемой областях.

Следует отметить, что успешное применение предложенной методологии НИР требует от сотрудников исследовательских подразделений расширять свои компетенции в смежных областях, связанных с технологиями разработки ПО и аппаратного обеспечения, анализом особенностей лицензионных соглашений, патентной информации и защиты объектов интеллектуальной собственности.

Литература

1. Global printer market: Market size, status and forecast to 2030. — Verified Market Research, 2022. 150 p.
2. Большая советская энциклопедия: в 30 т. — 3-е изд. — М.: Советская энциклопедия, 1969–1986.
3. *Липицу Н. В., Липицу К. И.* Методология научного исследования. — Краснодар: КубГАУ, 2013. 290 с.
4. ГОСТ Р 15.101-2021 Система разработки и постановки продукции на производство. Порядок выполнения научно-исследовательских работ. — М.: Российский институт стандартизации, 2021. 13 с.
5. *Гидроец М. О., Гришанова Л. И.* Методологии разработки программного продукта // Системный анализ и логистика, 2020. № 4. С. 45–53. doi: 10.31799/2007-5687-2020-4-45-53.
6. *Boehm B., Hansen W. J.* Spiral development: Experience, principles, and refinements. — Pittsburgh, PA, USA: Carnegie-Mellon University, Software Engineering Institute, 2000. 37 p.
7. *Маркова Н. А.* Перспективы ретивого производства и потребления программ // Системы и средства информатики, 2006. Т. 16. № 1. С. 277–288.
8. *Safonov I. V., Kurilin I. V., Rychagov M. N., Tolstaya E. V.* Adaptive image processing algorithms for printing. — Singapore: Springer Nature, 2018. 304 p.
9. *Safonov I. V., Kurilin I. V., Rychagov M. N., Tolstaya E. V.* Document image processing for scanning and printing. — Switzerland: Springer Nature, 2019. 306 p.
10. *Savoia A.* Pretotype it: Make sure you are building the right it before you build it right. — First Pretotype Edition, 2011. 71 p.

11. Virzi R. A., Sokolov J. L., Karis D. Usability problem identification using both low- and high-fidelity prototypes // SIGCHI Conference on Human Factors in Computing Systems Common Ground Proceedings. P. 236–243. doi: 10.1145/238386.238516.
12. Рудаков П. И., Сафонов И. В. Обработка сигналов и изображений. MatLab5x. — М.: Диалог-МИФИ, 2000. 416 с.

Поступила в редакцию 20.03.23

RESEARCH METHODOLOGY FOR CREATING FUNCTIONS OF SCANNING AND PRINTING SYSTEMS

I. V. Safonov¹ and I. A. Matveev²

¹National Research Nuclear University “MEPhI,” 31 Kashirskoe Shosse, Moscow 115409, Russian Federation

²Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: Scanning, copying, and printing devices are widely used in industry, government, and for personal needs. Combined with application and system software, these devices constitute scanning and printing systems. The formulation of a research methodology for creating the functions of such systems allows one formalizing and speeding up the development of new image processing methods. A review of the methodologies of scientific research, research and development in the industry, and software development has been done. The fact of implementation of the research result in the product was chosen as a success criterion. Based on a retrospective analysis, the factors that prevented the implementation of the results were identified. At the research stage, these factors can be considered as risks. Using the analytic hierarchy process, risks are ranked by their importance. A research methodology is proposed for creating new functions for scanning and printing systems based on risk analysis and management.

Keywords: research methodology; risk management; analytic hierarchy process

DOI: 10.14357/08696527230306

EDN: QEYMKA

References

1. Global printer market: Market size, status and forecast to 2030. 2022. Verified Market Research. 150 p.
2. *Bol'shaya sovetskaya entsiklopediya*: v 30 t. [Great Soviet Encyclopedia: in 30 vols]. 1969–1986. 3rd ed. Moscow: Sovetskaya entsiklopediya.
3. Lipchiu, N. V., and K. I. Lipchiu. 2013. *Metodologiya nauchnogo issledovaniya*. [Methodology of scientific research]. Krasnodar: KubGAU. 290 p.
4. GOST R 15.101-2021. 2021. Sistema razrabotki i postanovki produktii na proizvodstvo. Poryadok vypolneniya nauchno-issledovatel'skikh rabot [System of product development and launching into manufacture. Procedure of scientific researches and development]. Moscow: Russian Standardization Institute. 13 p.

5. Gidroyets, M. O., and L. I. Grishanova. 2020. Metodologii razrabotki programmogo produkta [Software development methodology]. *Sistemnyy analiz i logistika* [System Analysis and Logistics] 4(26):45–53. doi: 10.31799/2007-5687-2020-4-45-53.
6. Boehm, B., and W. J. Hansen. 2000. *Spiral development: Experience, principles, and refinements*. Pittsburgh, PA: Carnegie-Mellon University, Software Engineering Institute. 37 p.
7. Markova, N. A. 2006. Perspektivy retivogo proizvodstva i potrebleniya programm [Prospects of agile software development and consumption]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16(1):277–288.
8. Safonov, I. V., I. V. Kurilin, M. N. Rychagov, and E. V. Tolstaya. 2018. *Adaptive image processing algorithms for printing*. Singapore: Springer Nature. 304 p.
9. Safonov, I. V., I. V. Kurilin, M. N. Rychagov, and E. V. Tolstaya. 2019. *Document image processing for scanning and printing*. Switzerland: Springer Nature. 306 p.
10. Savoia, A. 2011. *Prototype it: Make sure you are building the right it before you build it right*. First Pretotype Edition. 71 p.
11. Virzi, R. A., J. L. Sokolov, and D. Karis. 1996. Usability problem identification using both low-and high-fidelity prototypes. *SIGCHI Conference on Human Factors in Computing Systems Common Ground Proceedings*. 236–243. doi: 10.1145/238386.238516.
12. Rudakov, P. I., and I. V. Safonov. 2000. *Obrabotka signalov i izobrazheniy. MatLab5x* [Processing of signals and images. MatLab5x]. Moscow: Dialog-MEPHI. 416 p.

Received March 20, 2023

Contributors

Safonov Ilia V. (b. 1971) — Candidate of Science (PhD) in technology, associate professor, National Research Nuclear University “MEPhI,” 31 Kashirskoe Shosse, Moscow 115409, Russian Federation; ilia.safonov@gmail.com

Matveev Ivan A. (b. 1973) — Doctor of Science in technology, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; matveev@frccsc.ru

ПРОБЛЕМЫ ФОРМИРОВАНИЯ СИСТЕМЫ СОГЛАСОВАННЫХ ЛОКАЛЬНЫХ ГЕОГРАФИЧЕСКИХ ОНТОЛОГИЙ

*Д. А. Никишин*¹

Аннотация: Рассматривается сущность геоонтологии как средства представления географического знания и решаемые ею задачи. Представлена многоуровневая система локальных геоонтологий, рассматриваемая как база для перспективных географических информационных систем (ГИС). Выделены проблемы реализации такой системы: выработка системы конвенционально понимаемых сущностей (геоконцептов), свойств и классификаций, основанной на согласованном подходе к структурированию и формализации моделируемых явлений; обеспечение пространственной и логической связанности компонентов модели местности; обеспечение взаимной согласованности локальных онтологий (интерфейсов взаимодействия) в аспектах генерализации и вариантности, а также, при необходимости, альтернативности и темпоральности для решения задач на «стыке» локальных геоонтологий.

Ключевые слова: геоонтологии; геоинформационные системы; моделирование географических явлений; система локальных геоонтологий; динамические геоонтологии

DOI: 10.14357/08696527230307

EDN: QKXGUK

1 Введение

В современных условиях возникли возможности внедрения ряда парадигм для организации географических данных (геоданных), которые желательно учитывать при разработке перспективных ГИС [1]. К таким парадигмам можно отнести [2]: обеспечение пространственной и логической связанности модели местности; создание целостной системы локальных геоонтологий [3]; поддержку множественных представлений геообъектов и местности в целом в аспектах генерализации, контекстной вариантности и альтернативности [4]; темпоральное моделирование [5]; трехмерное (3D) моделирование. Также нужно отметить актуальность создания надежных систем для автоматического географического вывода [2], включая задачу автоматического обеспечения целостности гео модели, что связано с внедрением пространственно-логической связанности геоданных.

Задачей данной публикации ставится рассмотрение парадигмы геоонтологии, ее задач и выявление проблем развития геоонтологий как целостной системы локальных геоонтологий [3].

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, dmnikishin@mail.ru

2 Сущность парадигмы информационной онтологии

Термин «онтология» заимствован из философии, где означал систематическое описание всего сущего [6]. Термин «информационная онтология» впервые введен в область науки и техники в 1984 г., а его концепция в целом была определена к 1991 г. Примерно в то же время произошло внедрение концепции онтологии в геоинформатику, и в последующие десятилетия многие исследования в области геоинформации были сосредоточены на теме геоонтологий [7]. Онтологии в настоящее время рассматриваются как наиболее перспективное и эффективное средство представления знаний [8].

Онтология в информатике, согласно современным толкованиям, является «точной спецификацией концептуализации предметной области», которая должна содержать тезаурус терминов и определенные спецификации. Онтологии содержат взаимосвязанные факты, объединенные в классы объектов и отношений, соответствующие выбранной проблемной области. По сути, онтология — это наиболее полная концептуализация проблемной области [3, 9].

Обобщая определения, приведенные в [6, 8, 10–12] и ряде других источников, можно уточнить содержание информационной онтологии как целостной системы формализованных знаний, содержащей определения свойств и понятий (концептов, классов) предметной области, объединенных отношениями между ними. «Свойство — категория, выражающая такую сторону объекта (сущности), которая обуславливает его различие или общность с другими объектами и обнаруживает себя при сопоставлении различных объектов. Объекты являются не столько носителями свойств, сколько их совокупностью, что позволяет считать свойства объектообразующей категорией» [13]. Обобщение моделей объектов посредством понятий (классов) обусловлено необходимостью сопоставления и обобщения схожих сущностей. Это обеспечивает концептуальную основу для общей характеристики явлений. Отношение отражает пространственную или логическую взаимосвязь между географическими понятиями или конкретными геообъектами [7].

3 Задачи информационной онтологии

Если «цель философской онтологии состоит в том, чтобы лучше понять (и объяснить) структуру мира» [7], то целями онтологии информационной можно считать следующее.

1. На основе онтологий осуществляется моделирование предметной области, которое предполагает концептуализацию (абстракцию и генерализацию) явлений предметной области моделями отдельных объектов. Онтология должна обеспечивать последовательное и полное структурирование предметной области [8].
2. Еще одна роль онтологии — определение общего словаря для пользователей (людей, компьютеров или программных агентов), которым необходимо обме-

ниваться информацией об определенной предметной области [6, 7]. Для этого содержащиеся в онтологии концепты и отношения должны быть приняты и разделены использующим ее сообществом [11]. По сути, онтология служит конвенциональной основой для согласования понимания пользователями сущностей предметной области. В этом качестве онтология может служить основой для обмена между различными источниками информации [8, 11].

3. Онтология призвана обеспечивать понимание семантики данных не только человеком, но и компьютером, что позволяет автоматизировать запросы данных, логические рассуждения и вывод новых, неявных знаний из имеющихся данных, но требует предельно четких, формализованных и стандартизованных определений понятий и отношений между ними [7, 10, 11].
4. Онтология позволяет отделять знание о предметной области от реализации конкретных информационных систем. Это дает возможность мигрировать с одной информационной системы на другую, а также адаптироваться к изменениям знания [7].

Таким образом, информационная онтология служит основой для приобретения и совершенствования знания о предметной области, а также для передачи этого знания — т. е., по сути, выполняет ту же самую задачу, что и онтология философская. Но, в отличие от последней, у информационной онтологии есть два существенных отличия. Во-первых, она обычно создается целенаправленно, в объеме, необходимом и достаточном для решения конкретных, насущных задач моделирования. Соответственно, именно решаемые задачи определяют параметры создаваемой онтологии, а основным критерием служит ее применимость, а не ее полнота. Другая особенность онтологии в информатике заключается в наличии четких, формальных определений, представленных аксиоматически или эксплицитно и раскрывающих содержание понятий (концептов) и категорий свойств [6], а также правила и ограничения их применимости [8].

4 Проблемы подхода на основе системы локальных онтологий

Исследования последних лет показывают, что создание всеобъемлющей онтологии можно считать утопией [12]. Это обусловлено как необъятностью круга возможных явлений и неограниченностью их детализации, так и целым рядом причин, рассмотренных ниже. В связи с этим в последнее время получила распространение идея локальных (или модульных) онтологий, адаптированных к конкретным потребностям их использования [14, 15]. В [7, 8, 10] упоминаются несколько уровней подобного разделения онтологий.

1. Онтология верхнего уровня определяет общие, универсальные для многих предметных областей понятия и отношения между ними и выступает как основа для применения локальных онтологий.
2. Локальные (специализированные) онтологии в современных исследованиях подразделяются на следующие уровни:

- (а) онтология, представляющая специфичную терминологическую базу для конкретной предметной области. Принятые в этой области термины получаются либо специализацией терминов более общей онтологии, либо путем введения новых специфических терминов. Соответственно, происходит спецификация ограничений, накладываемых конкретной предметной областью¹ (см., например, [8]);
- (б) онтология для конкретной задачи;
- (в) онтология приложения, описывающая концепты и отношения для конкретной реализации (информационной системы, базы данных и т. п.).

Вместе с тем у подхода на основе локальных онтологий имеется ряд проблем.

1. Одна из проблем заключается в озабоченности по поводу фрагментации знаний, присущих распространению неопределенного числа микроонтологий в условиях отсутствия общего ядра. Это особенно важно в случае, когда несколько моделей создаются в сильно разрозненных контекстах наблюдения и интерпретации².
2. Даже в рамках одной предметной области могут существовать варианты онтологий, отличные по масштабу, контексту использования или представляющие альтернативные взгляды на одну и ту же предметную область. Аспект генерализации наглядно представлен масштабным рядом условных обозначений топографических и других карт [16]. Вариантность вызвана условиями конкретной задачи [2, 4]³; примером могут служить онтологии топографической и морской картографии. Под альтернативностью здесь понимается существование различающихся онтологий при условии, что речь идет об одной предметной области и об одном контексте использования, при этом «они должны выдерживать критические испытания при столкновении с реальностью» [17]. Примером могут служить отечественная и западная традиции картографии, топологическая (OpenGIS) или нетопологическая (GeoRSS — Geographically encoded objects for Really Simple Syndication) модели геометрии геообъектов.
3. Для решения задач на стыке предметных областей может потребоваться совместное использование локальных онтологий.

¹ В разных локальных онтологиях могут присутствовать одноименные, но разные (в том числе несовместимые) по сущности термины. Если же термины в разных областях схожи, то, скорее всего, они образуют универсальное понятие, которое может быть отнесено к общей онтологии.

² Вообще, вопрос о разделении предметных областей не так прост. Например, относительно геоонтологии возникает дилемма: это онтология предметной области [8] или она должна быть отнесена к уровню базовой онтологии? Ведь очевидно, что во многих различных предметных областях нужна концепция для выражения географического пространства и расположения [7].

³ Так, для архитектора, инженера, строителя, экономиста взгляд на строящийся объект определен теми задачами, которые каждый из них решает в отношении этого здания. Соответственно, их деятельность будет базироваться на достаточно близких, но тем не менее различающихся онтологиях.

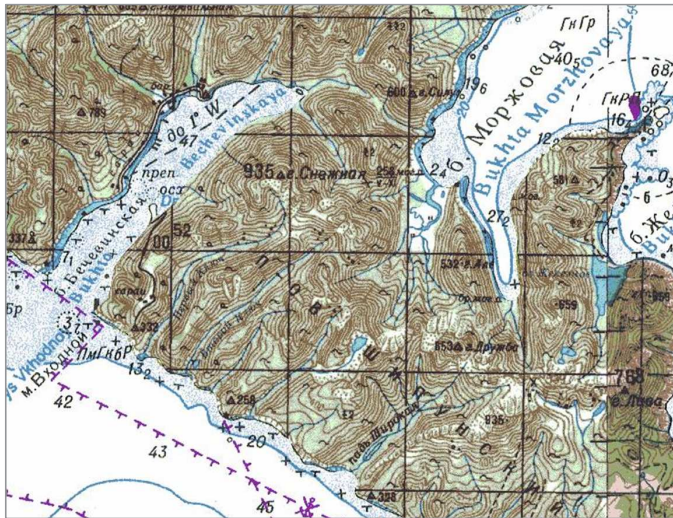


Рис. 1 Пример различного положения береговой линии при совмещении топографической и морской навигационной карт [18]

Примером «стыка» онтологий одного уровня детализации может служить решение задач «суша–море», предполагающих совместное использование геоделей суши и прибрежного шельфа. Однако исторически сложившееся ведомственное разделение сфер моделирования суши и моря привело к различиям в подходах к представлению соответствующих геоданных. Для их совместного использования требуется разрешить ряд проблем, в том числе согласования используемых ими онтологий. И если проблемы унификации семантики их сущностей решаются достаточно тривиально, то проблема согласования пространственного положения геоделей суши и моря (в плане и по высоте) пока не имеет однозначного решения ввиду отсутствия количественной связи между используемыми ими системами отсчета координат топографических высот и навигационных глубин [18]. Так, в результате совмещения топографической и морской навигационной карт наблюдается сдвиг береговой линии (рис. 1), что обусловлено особенностями выбора их уровневой поверхности.

Примером «стыка» онтологий с различными уровнями детализации может служить интеграция онтологии топографической модели крупного масштаба и онтологии моделей BIM (Building Information Model) уровня отдельного инженерного сооружения [19], например при решении градостроительных задач (рис. 2). Факторами, затрудняющими их совместное применение помимо различий в семантике и используемых системах координат здесь выступают необходимость согласования моделей по масштабу (согласования детальности этих моделей), а также, при необходимости, согласование элементов геопространственных моделей различной размерности (2D, 3D, 4D).



Рис. 2 Иллюстрация совместного использования разномасштабных моделей на примере проекта застройки из [19]

Итак, для решения подобных задач необходимо обеспечить «бесшовную» увязку (согласование) соответствующих локальных онтологий. При этом каждая из этих онтологий должна учитывать аспекты, смежные с другими онтологиями, с тем чтобы они были совместимы.

Также нужно отметить, что структура онтологии и отдельные понятия могут быть подвержены изменениям во времени [8, 17], «при необходимости онтология должна быть улучшена и расширена» [10]. Это обуславливает динамику онтологии, пересмотр геознания в контексте его развития, что позволяет выделить два момента. Во-первых, при формировании онтологии нужно иметь в виду не только ее изначальное построение, но и обеспечение ее ведения. Во-вторых, представляется перспективной задачей разработка **динамических геоонтологий**, допускающих не просто их периодическую актуализацию, но и обеспечивающих их темпоральность. Это, в свою очередь, предполагает, что для решения задачи на определенный момент времени необходимо обеспечение соответствия данных и используемой ими версии геоонтологии. В этом же контексте нужно отметить проблему решения задач, относящихся к периодам времени, лежащим на «стыке» двух разновременных версий геоонтологии и соответствующих им структур геоданных.

5 Заключение

В настоящее время признан огромный потенциал и перспективы применения онтологий как в области географических, так и информационных технологий (в том числе искусственного интеллекта) [7]. И можно было бы ожидать, что основные географические концепции уже представлены в современных географических онтологиях, тем не менее до сих пор не существует всеобъемлющей и согласован-

ной геоонтологии [8]. Достижения в географических онтологиях по-прежнему сосредоточены на теоретических исследованиях; спроектированы и реализованы лишь несколько небольших экспериментальных систем, но пока еще не появился крупный проект географической онтологии [7].

Реализация масштабного проекта геоонтологии представляется возможной на основе системы локальных геоонтологий. Но для интеграции локальных геоонтологий как системы представляется необходимым решение следующих проблем:

- выработка системы конвенционально понимаемых сущностей (геоконцептов), свойств и классификаций на основе общепринятой, базовой концепции; при этом необходим согласованный подход к структурированию и формализации моделируемых явлений, в том числе с целью обеспечения возможности автоматического географического вывода;
- обеспечение пространственной и логической связанности компонентов модели местности;
- обеспечение взаимной согласованности («бесшовной» увязки) локальных онтологий и связываемых с ними наборов данных, определение интерфейсов взаимодействия в аспектах генерализации и вариантности, а также, при необходимости, альтернативности и темпоральности.

Литература

1. *Никишин Д. А.* Направления развития методологической базы для работы с гео-данными в перспективных геоинформационных системах // Системы и средства информатики, 2023. Т. 33. № 2. С. 34–45. doi: 10.14357/08696527230204.
2. *Laurini R.* Some philosophical issues regarding geometric modeling for geographic information and knowledge systems // The philosophy of GIS / Ed. T. Tambassi. — Springer geography ser. — Cham: Springer, 2019. P. 25–50. doi: 10.1007/978-3-030-16829-2_2.
3. *Дулин С. К., Никишин Д. А.* Подходы к интеграции прикладных концептуальных схем в составе унифицированной геоонтологии // Системы и средства информатики, 2020. Т. 30. № 2. С. 68–77. doi: 10.14357/08696527200207.
4. *Никишин Д. А.* Подход к разработке концептуальной схемы базы геоданных с поддержкой многовариантного представления геообъектов // Системы и средства информатики, 2022. Т. 32. № 2. С. 128–136. doi: 10.14357/08696527220212.
5. *Никишин Д. А.* Обзор подходов к пространственно-временному моделированию и выявление основных тенденций развития T-GIS // Системы и средства информатики, 2022. Т. 32. № 3. С. 50–62. doi: 10.14357/08696527220305.
6. *Gruber T. R.* A translation approach to portable ontology specifications // Knowl. Acquis., 1993. Vol. 5. Iss. 2. P. 199–220. doi: 10.1006/knac.1993.1008.
7. *Li H.* Geographic ontology // Advances in cartography and geographic information engineering / Eds. J. Wang and F. Wu. — Singapore: Springer, 2021. P. 479–501. doi: 10.1007/978-981-16-0614-4_13.
8. *Kavouras M., Kokla M., Tomai E.* Semantically-aware systems: Extraction of geosemantics, ontology engineering and ontology integration // Geographic hypermedia /

- Eds. E. Stefanakis, M. P. Peterson, C. Armenakis, V. Delis. — Lecture notes in geoinformation and cartography ser. — Berlin–Heidelberg: Springer, 2006. P. 257–273. doi: 10.1007/978-3-540-34238-0_14.
9. *McDaniel M., Storey V. C.* Evaluating domain ontologies: Clarification, classification, and challenges // *ACM Comput. Surv.*, 2019. Vol. 52. Iss. 4. Art. 70. 44 p. doi: 10.1145/3329124.
 10. *Wang Y., Gong J., Wu X.* Geospatial semantic interoperability based on ontology // *Geo-spatial Information Science*, 2007. Vol. 10. Iss. 3. P. 204–207. doi: 10.1007/s11806-007-0071-7.
 11. *Zhang C., Zhao T., Li W.* Conceptual frameworks of geospatial semantic web // *Geospatial semantic web*. — Cham: Springer, 2015. P. 35–56. doi: 10.1007/978-3-319-17801-1_2.
 12. *Couclelis H.* Unpacking the “I” in GIS: Information, ontology, and the geographic world // *The philosophy of GIS* / Ed. T. Tambassi. — Springer geography ser. — Cham: Springer, 2019. P. 3–24. doi: 10.1007/978-3-030-16829-2_1.
 13. *Кравченко Ю. А.* Основы конструирования систем геомоделирования. Кн. 2. Информационное геомоделирование: модели и методы. — Новосибирск: СГГА, 2008. Ч. 2. 316 с.
 14. *Gangemi A.* Ontology design patterns for semantic web content // *The semantic web* / Eds. Y. Gi, E. Motta, V. Benjamins, M. Musen. — Lecture notes in computer science ser. — Berlin–Heidelberg: Springer, 2005. Vol. 3729. P. 262–276. doi: 10.1007/11574620_21.
 15. *Gangemi A., Presutti V.* Towards a pattern science for the semantic web // *Semant. Web*, 2010. Vol. 1. No. 1-2. P. 61–68.
 16. *Никитин Д. А.* Процессы генерализации в аналоговой и цифровой картографии // *Системы и средства информатики*, 2018. Т. 28. № 3. С. 204–216. doi: 10.14357/08696527180316.
 17. *Grenon P., Smith B.* SNAP and SPAN: Towards dynamic spatial ontology // *Spat. Cogn. Comput.*, 2004. Vol. 4. Iss. 1. P. 69–104. doi: 10.1207/s15427633scc0401_5.
 18. *Жуков Ю. Н.* Проблемы согласования навигационной и топографической геоинформации // *Системы и средства информатики*, 2008. Спец. вып.: Геоинформационные технологии. С. 273–281.
 19. *Скворцов А. В.* Модели данных ВМ для инфраструктуры // *САПР и ГИС автомобильных дорог*, 2015. № 1(4). С. 16–23. doi: 10.17273/CADGIS.2015.1.2.

Поступила в редакцию 14.08.23

PROBLEMS OF FORMATION OF A SYSTEM OF COORDINATED LOCAL GEOGRAPHICAL ONTOLOGIES

D. A. Nikishin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The essence of geontology as a means of representing geographical knowledge and the tasks it solves are considered. A multilevel system of local

geontologies is presented considered as a base for promising geoinformation systems. The problems of implementing such a system are highlighted: the development of a system of conventionally understood entities (geoconcepts), properties and classifications based on a coordinated approach to structuring and formalizing the simulated phenomena; ensuring spatial and logical connectivity of the components of the terrain model; ensuring mutual consistency of local ontologies (interaction interfaces) in the aspects of generalization and variation as well as, if necessary, alternativeness and temporality to solve the problems at the “junction” of local geontologies.

Keywords: gerontology; geoinformation systems; modeling of geographical phenomena; system of local gerontology; dynamic geontology

DOI: 10.14357/08696527230307

EDN: QKXGUK

References

1. Nikishin, D. A. 2023. Napravleniya razvitiya metodologicheskoy bazy dlya raboty s geodannymi v perspektivnykh geoinformatsionnykh sistemakh [Directions of development of the methodological base for working with geodata in perspective geoinformation systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):34–45. doi: 10.14357/08696527230204.
2. Laurini, R. 2019. Some philosophical issues regarding geometric modeling for geographic information and knowledge systems. *The philosophy of GIS*. Ed. T. Tambassi. Springer geography ser. Cham: Springer. 25–50. doi: 10.1007/978-3-030-16829-2_2.
3. Dulin, S. K., and D. A. Nikishin. 2020. Podkhody k integratsii prikladnykh kontseptual'nykh skhem v sostave unifitsirovannoy geoontologii [Approaches to the integration of the application conceptual schemas in the unified geontology]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(2):68–77. doi: 10.14357/08696527200207.
4. Nikishin, D. A. 2022. Podkhod k razrabotke kontseptual'noy skhemy bazy geodannykh s podderzhkoy mnogovariantnogo predstavleniya geoob'ektov [An approach to the development of a conceptual scheme of a geodata database with support for a multivariate representation of geobjects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):128–136. doi: 10.14357/08696527220212.
5. Nikishin, D. A. 2022. Obzor podkhodov k prostranstvenno-vremennomu modelirovaniyu i vyyavlenie osnovnykh tendentsiy razvitiya T-GIS [Overview of approaches to space–time modeling and the main trends in the development of T-GIS]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):50–62. doi: 10.14357/08696527220305.
6. Gruber, T. R. 1993. A translation approach to portable ontology specifications. *Knowl. Acquis.* 5(2):199–220. doi: 10.1006/knac.1993.1008.
7. Li, H. 2021. Geographic ontology. *Advances in cartography and geographic information engineering*. Eds. J. Wang and F. Wu. Singapore: Springer. 479–501. doi: 10.1007/978-981-16-0614-4_13.
8. Kavouras, M., M. Kokla, and E. Tomai. 2006. Semantically-aware systems: Extraction of geosemantics, ontology engineering, and ontology integration. *Geographic hypermedia*. Eds. E. Stefanakis, M. P. Peterson, C. Armenakis, and V. Delis. Lecture notes

- in geoinformation and cartography ser. Berlin–Heidelberg: Springer. 257–273. doi: 10.1007/978-3-540-34238-0_14.
9. McDaniel, M., and V. C. Storey. 2019. Evaluating domain ontologies: Clarification, classification, and challenges. *ACM Comput. Surv.* 52(4):70. 44 p. doi: 10.1145/3329124.
 10. Wang, Y., J. Gong, and X. Wu. 2007. Geospatial semantic interoperability based on ontology. *Geo-spatial Information Science* 10(3):204–207. doi: 10.1007/s11806-007-0071-7.
 11. Zhang, C., T. Zhao, and W. Li. 2015. Conceptual frameworks of geospatial semantic web. *Geospatial semantic web*. Cham: Springer. 35–56. doi: 10.1007/978-3-319-17801-1_2.
 12. Couclelis, H. 2019. Unpacking the “I” in GIS: Information, ontology, and the geographic world. *The philosophy of GIS*. Ed. T. Tambassi. Springer geography ser. Cham: Springer. 3–24. doi: 10.1007/978-3-030-16829-2_1.
 13. Kravchenko, Yu. A. 2008. *Osnovy konstruirovaniya sistem geomodelirovaniya. Kn. 2. Informatsionnoe geomodelirovanie: modeli i metody* [Fundamentals of designing geomodeling systems. Book 2. Information geomodeling: Models and methods]. Novosibirsk: SGGa. Part 2. 316 p.
 14. Gangemi, A. 2005. Ontology design patterns for semantic web content. *The semantic web*. Eds. Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen. Lecture notes in computer science ser. Berlin–Heidelberg: Springer. 3729:262–276. doi: 10.1007/11574620_21.
 15. Gangemi, A., and V. Presutti. 2010. Towards a pattern science for the semantic web. *Semant. Web* 1(1-2):61–68.
 16. Nikishin, D. A. 2018. Protsessy generalizatsii v analogovoy i tsifrovoy kartografii [Processes of generalization in analog and digital cartography]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):204–216. doi: 10.14357/08696527180316.
 17. Grenon, P., and B. Smith. 2004. SNAP and SPAN: Towards dynamic spatial ontology. *Spat. Cogn. Comput.* 4(1):69–104. doi: 10.1207/s15427633scc0401_5.
 18. Zhukov, Yu. N. 2008. Problemy soglasovaniya navigatsionnoy i topograficheskoy geoinformatsii [Problems of coordination of navigation and topographic geoinformation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Spets. vyp.: Geoinformatsionnye tekhnologii [Special issue: Geoinformation technologies] 273–281.
 19. Skvortsov, A. V. 2015. Modeli dannykh BIM dlya infrastruktury [BIM data models for infrastructure]. *SAPR i GIS avtomobil'nykh dorog* [CAD & GIS for Roads] 1(4):16–23. doi: 10.17273/CADGIS.2015.1.2.

Received August 14, 2023

Contributor

Nikishin Dmitry A. (b. 1976)— Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; dmnikishin@mail.ru

НЕКОТОРЫЕ ПОДХОДЫ К АНАЛИЗУ ФАКТОРОВ, ВЛИЯЮЩИХ НА ИНФОРМАЦИОННУЮ БЕЗОПАСНОСТЬ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

А. А. Зацаринный¹, А. П. Сучков²

Аннотация: Рассматривается проблематика обеспечения необходимого уровня информационной безопасности (ИБ) систем искусственного интеллекта (СИИ), что становится одним из ключевых условий их внедрения в государственную практику. Отмечается, что основную специфическую угрозу несут процедуры обучения СИИ, в ходе которых настраивается и может меняться не только информационная база, но и алгоритмы интеллектуальных выводов. Предложена технология выявления особенностей обеспечения ИБ СИИ, позволяющая охарактеризовать совокупность существующих угроз. Обсуждены специфические вопросы защиты информации СИИ и ряд мероприятий по обеспечению ИБ этих систем.

Ключевые слова: информационная безопасность; системы искусственного интеллекта; процедуры обучения; виды защиты информации

DOI: 10.14357/08696527230308

EDN: QNXZBN

1 Введение

Принятие Национальной стратегии развития искусственного интеллекта (ИИ) в России на период до 2030 г. (утверждена Указом Президента РФ от 10.10.2019 № 490) вызвало целый комплекс программных и практических мероприятий по разработке и внедрению СИИ в различных сферах деятельности. Вместе с тем использование СИИ наряду с очевидными преимуществами порождает новые угрозы ИБ. Это обстоятельство требует особого внимания к анализу факторов, определяющих функционирование СИИ в соответствии с требованиями по назначению. По существу, внедрение СИИ приводит к принципиальному изменению классической парадигмы защиты информации, а именно: на первый план, в отличие от традиционного примата конфиденциальности, выходят целостность и доступность, которые определяются корректностью реализации алгоритмов функционирования программным обеспечением, а также их полнотой и непротиворечивостью.

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ASuchkov@ipiran.ru

В связи с этим обеспечение необходимого уровня ИБ СИИ становится одним из ключевых условий их внедрения в информационные, прежде всего государственные, системы. Именно это обуславливает актуальность уточнения и дополнения действующих нормативных документов, определяющих формирование требований к ИБ информационных систем с учетом особенностей технологий ИИ (ТИИ).

Отметим существующие базовые определения (ГОСТ Р 50922-2006).

Информационная безопасность — это защищенность информации и поддерживающей инфраструктуры от случайных или преднамеренных воздействий естественного или искусственного характера, чреватых нанесением ущерба владельцам или пользователям информации и поддерживающей инфраструктуры.

Защита информации — это совокупность мероприятий, направленных на обеспечение конфиденциальности и целостности обрабатываемой информации, а также доступности информации для пользователей.

Объект защиты информации — информация, или носитель информации, или информационный процесс, которые необходимо защищать в соответствии с целью защиты информации.

Целью защиты информации ставится сведение к минимуму потерь в управлении, вызванных нарушением целостности данных, их конфиденциальности или недоступности информации для потребителей.

Вопросы существующего нормативно-правового обеспечения подробно освещены в [1, 2]. Отмечается, что преимущественно оно ориентировано на защиту конфиденциальности информации, при этом обеспечение целостности и доступности информации носит вспомогательный характер. Такой подход приводит к возможному ущербу *функциональной безопасности* автоматизированной системы как способности корректно и эффективно выполнять заданные функциональные задачи, в том числе в условиях возможного противодействия потенциального нарушителя [2]. Там же отмечается, что важную роль играет обеспечение *системной безопасности*, связанной с предотвращением системных ошибок в реализации организационно-технических мер защиты, связанных с человеческим фактором на этапе эксплуатации автоматизированных систем, а также системные уязвимости, заложенные на этапах проектирования.

В связи с этим в [3] сделан вывод о необходимости корректировки нормативно-методической базы с ориентацией на широкое использование методов мониторинга событий в цифровом пространстве с учетом среди прочего угроз функциональной и системной безопасности.

Исходя из нормативных документов, для изучения вопроса ИБ СИИ необходимо проанализировать объекты защиты информации — собственно информацию (на предмет искажения, изменения, исследования, записи или уничтожения информации) и информационные процессы (на предмет возможности появления недекларируемых возможностей (НДВ) и влияния на сертификацию СИИ).

Одна из основных проблем обеспечения ИБ СИИ на основе технологии машинного обучения заключается в том, что создание и функционирование таких

систем невозможно без разнообразных процедур обучения, дообучения и самообучения с использованием зачастую огромных массивов данных из различных источников — баз данных, баз знаний, экспертных данных, коллекций изображений и аудиозаписей, сред сенсорного и имитационного моделирования, «цифровых двойников». Очень часто источники этих данных не входят в число доверенных и формируются в открытом доступе с целью тестирования новых подходов к созданию СИИ самыми разными авторами и научным сообществом в целом. Таким образом, с точки зрения ИБ СИИ необходимо изучить этот объект защиты.

Другая специфическая особенность — фактор применения интеллектуальных автономных и мобильных СИИ, что предъявляет специфические требования к сохранению конфиденциальности информации и информационных процессов в смысле расширения спектра угроз ИБ (возможность извлечения и инверсии моделей и т. п.).

С точки зрения информационных процессов, влияющих на ИБ СИИ, к одним из основных специфических угроз ИБ ИИ следует отнести процедуры обучения СИИ, в ходе которых настраиваются и могут меняться не только информационная база, но и алгоритмы интеллектуальных выводов. Проблема состоит в том, что даже небольшие изменения в обучающих наборах данных могут привести к кардинальным изменениям интеллектуальных способностей таких систем, вплоть до их существенного нарушения. Важно также очень трудно проверяемое качество обучающих данных (репрезентативность, статистическая несмещенность, достаточный объем).

На рис. 1. приведен пример такого критичного изменения в обучающей выборке [4]. На рис. 1, *a* — изображение из проекта ImageNet по созда-

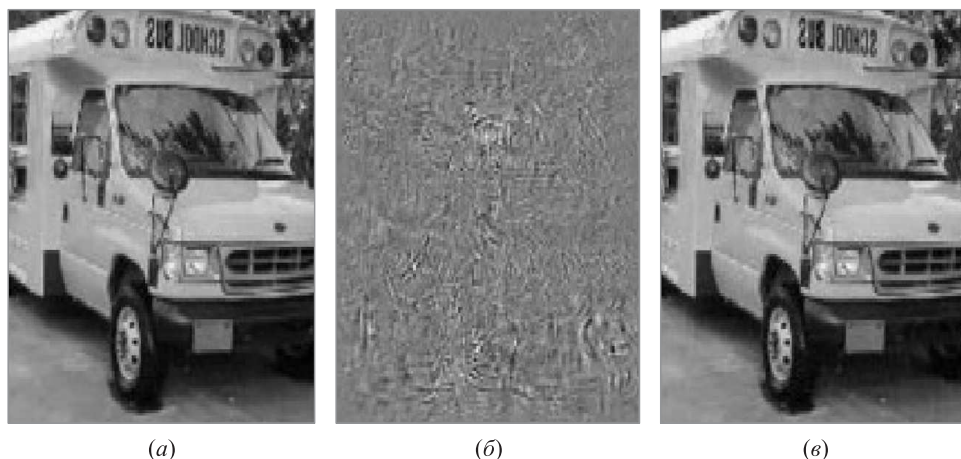


Рис. 1 Пример критичного изменения в обучающей выборке

нию и сопровождению массивной базы данных аннотированных изображений, предназначенного для отработки и тестирования методов распознавания образов и машинного зрения, которое глубокая сверточная сеть классифицировала правильно. На рис. 1, в — немного измененное изображение (на рис. 1, б показаны различия; их много, но плотность ничтожна и не различима человеком), которое сеть уже не смогла правильно распознать. Обнаружено, что подобные «составительные» изменения можно подобрать для любого изображения из базы, а не только для избранных, что представляет существенную угрозу функциональной безопасности СИИ.

Другая существенная угроза — невозможность воспроизведения процедур обучения СИИ, обусловленная инициализацией случайным образом весов и смещений сети, использованием стохастического градиентного спуска, случайным перемешиванием обучающих данных и последующим их разбиением на мини-пакеты нужного размера.

Таким образом, анализ специфики вопросов ИБ СИИ чрезвычайно актуален. В статье обсуждается один из возможных подходов к систематизации факторов, которые необходимо учитывать при определении возможных угроз ИБ СИИ на основе системного анализа информационных процессов функционирования СИИ.

2 Подходы к выявлению угроз информационной безопасности для систем искусственного интеллекта

Прежде всего отметим, что СИИ могут быть отнесены к одному из видов автоматизированных информационных систем. Это позволяет применять к ним все существующие виды защиты информации (правовая, техническая, криптографическая, физическая, функциональная), регламентированные существующей нормативно-методической базой.

Рассматривая стадии жизненного цикла (ЖЦ) СИИ (замысел, концепция, разработка, производство, применение, поддержка), можно выделить ряд факторов, в том числе специфических для СИИ и влияющих на их ИБ:

- (1) методы ИИ;
- (2) ТИИ;
- (3) вид системы ИИ;
- (4) стадии ЖЦ СИИ;
- (5) информационные процессы;
- (6) внешняя информационная среда;
- (7) угрозы ИБ;
- (8) мероприятия по защите информации;

- (9) нормативно-методические документы, регламентирующие защиту информации.

Их последовательный анализ позволяет определить особенности обеспечения ИБ СИИ и охарактеризовать совокупность существующих угроз и мер противодействия им.

Рассмотрим один из путей такого анализа (рис. 2).

1. Выбор метода. Одна из особенностей исследований в области ИИ связана с большим разнообразием применяемых методов ИИ, требующих своих специфических подходов к обеспечению ИБ. Так как носителем одной из основных угроз ИБ СИИ выступают процессы взаимодействия на разных стадиях ЖЦ с внешней информационной средой, необходимо выделить методы ИИ, непосредственно связанные с такими процессами (см. рис. 2).

2. Выбор технологии. На основе методов ИИ формируются ТИИ, которые могут быть как унимодальными, так и гибридными. Например, гибридные ТИИ анализа данных для обработки информации и автоматического распознавания образов включают:

- технологии автоматического анализа и распознавания сигналов и изображений различной физической природы (оптические, акустические, радиолокационные и др.);
- технологии автоматической обработки, улучшения, распознавания, синтеза и анализа звуковой и речевой информации, включая технологии автоматического перевода;
- технологии автоматической семантической обработки и анализа коллекций документов, содержащих текстовую и графическую информацию;
- технологии автоматической обработки и анализа неантропоморфных данных.

Из всего многообразия выберем ТИИ, тесно связанные с процессами обучения и пополнения баз знаний (см. рис. 2).

3. Определение вида системы искусственного интеллекта. С использованием ТИИ создаются реальные образцы СИИ. Для определения вида СИИ за основу можно взять стандарт ГОСТ Р 59277-2020, где рассмотрены следующие критерии для классификации СИИ:

- по степени автономности;
- по степени автоматизации;
- по архитектурному принципу;
- по структуре и процессам обработки знаний:
 - по модели знаний;
 - по управлению знаниями;
 - по методу обучения;

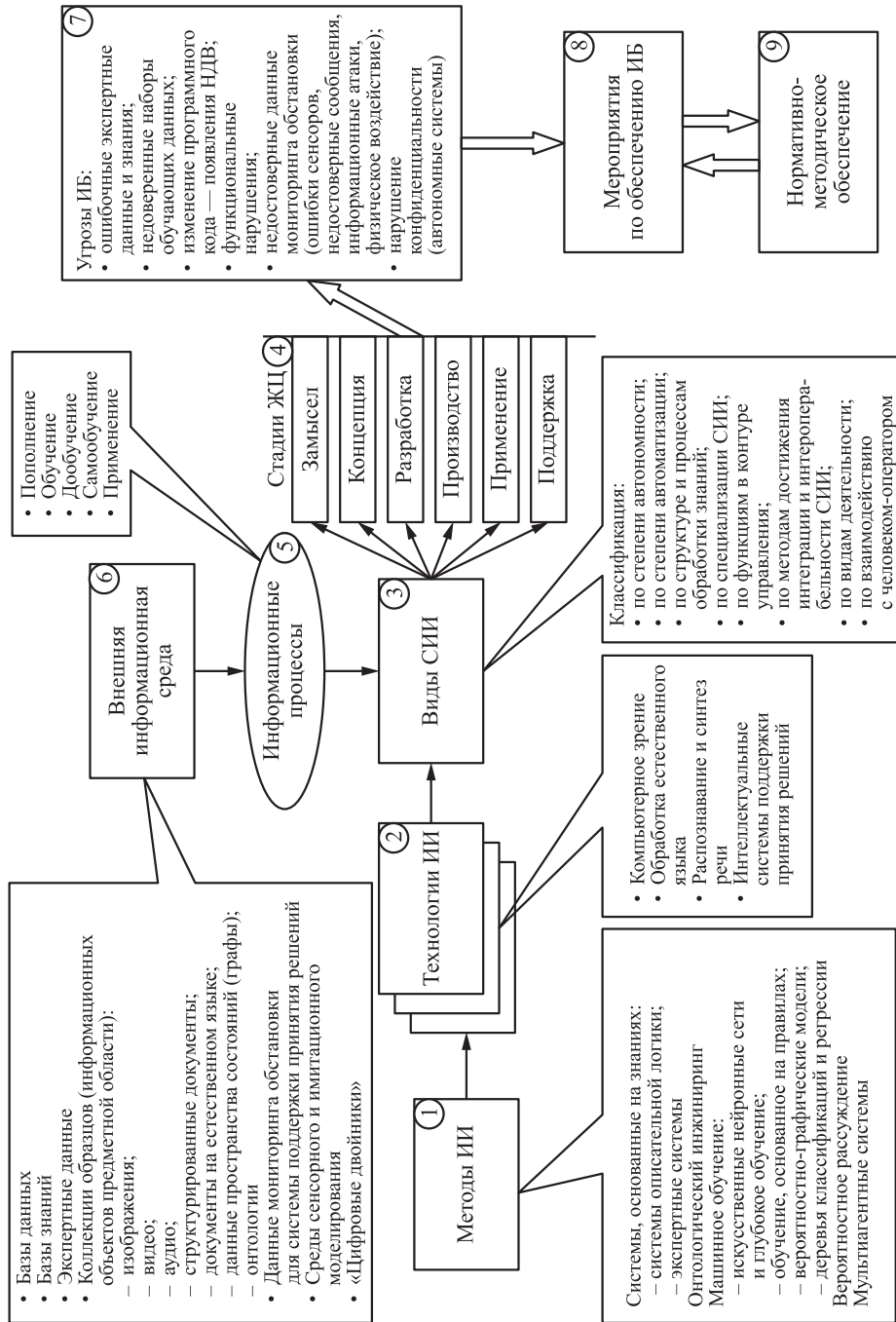


Рис. 2 Технология анализа специфики обеспечения ИБ СИИ

- по специализации СИИ:
 - специализированные (используют единый домен знаний);
 - комплексные (используют множество доменов знаний);
- по методам обработки информации;
- по функциям в контуре управления;
- по методам достижения интеграции и интероперабельности СИИ;
- по опасности последствий;
- по конфиденциальности;
- по видам деятельности;
- по взаимодействию с человеком-оператором.

На рис. 2 отображены релевантные поставленной задаче классы СИИ.

4. Определение угроз на рассматриваемой стадии жизненного цикла. На каждой из стадий ЖЦ СИИ (замысел, концепция, разработка, производство, применение по назначению, поддержка) можно выделить свой специфический набор угроз ИБ СИИ, связанный с процессами информационного взаимодействия (см. таблицу). В качестве примера общего анализа угроз можно привести [5].

Обеспечение безопасности информации СИИ складывается:

- из системной безопасности, связанной с корректностью формирования системных требований к СИИ на стадиях замысла и концепции, а также с корректным применением существующих средств и технологий защиты информации с учетом реальных условий их эксплуатации и угроз организационно-технического характера на стадиях разработки и применения [1];
- из технической защиты информации, блокирующей (предотвращающей) деструктивные действия потенциальных нарушителей на стадиях разработки, производства и поддержки;
- из функциональной безопасности, связанной с полной реализацией функций СИИ при условии реализации мероприятий по защите информации на всех стадиях ЖЦ [6], включая случайный риск, обусловленный неполнотой обучающих данных (см., например, [3]), а также зачастую невозможностью верифицировать процессы обучения в ходе тестирования СИИ (стадии разработки, производства и поддержки).

3 Подходы к защите информации для систем искусственного интеллекта

К основным видам защиты информации для СИИ можно отнести (ГОСТ Р 50922-2006):

- техническую защиту информации;
- криптографическую защиту информации, включая обучающие данные;

Примеры угроз ИБ СИИ на разных стадиях ЖЦ

Стадия ЖЦ	Угроза	Нарушения
Замысел	Некорректный выбор методов ИИ и принципов информационного взаимодействия	Системная безопасность
Концепция	Некорректные системные требования к ИБ информационного взаимодействия	Системная безопасность
Разработка	Невозможность реализации системных требований к ИБ	Системная безопасность Техническая защита и функциональная полнота: – некорректный выбор ТИИ; – недоверенные наборы обучающих данных
Производство	Невозможность реализации проектных решений по ИБ информационных взаимодействий	Техническая защита и функциональная полнота: – нарушение конфиденциальности; – проблемы тестирования; – проблема воспроизводимости
Применение по назначению	Неполная реализация системных требований по ИБ, ограничения функциональности	Техническая защита и функциональная полнота: – снижение декларируемых возможностей; – снижение устойчивости СИИ; – нарушение конфиденциальности
Поддержка	Некорректное дообучение и самообучение* Изменение программного кода	Техническая защита и функциональная полнота: – появление НДВ; – снижение декларируемых возможностей

* Виды информации для дообучения и возможные угрозы на стадии поддержки приведены на рис. 2.

- физическую защиту информации;
- правовую защиту информации.

Также, с точки зрения авторов, к основным видам защиты СИИ необходимо добавить:

- обеспечение системной и функциональной безопасности;
- документирование уровней готовности технологий;
- повышение устойчивости СИИ.

В рамках обеспечения системной безопасности необходимо определить целевые области применения СИИ и ее критические элементы, а также сформировать

реализуемые системные требования на основе анализа промышленных и технологических рисков.

Техническая и физическая защита информации осуществляется на основе существующих нормативно-методических документов и не отличается от комплекса мероприятий по обеспечению безопасности информации для автоматизированных информационных систем.

Вопросы криптографической защиты для СИИ могут иметь свою специфику. Так, в проекте ГОСТ [7] рассматривается *защищенное исполнение в нейросетевых СИИ процедуры (алгоритмов) классификации* — специальный режим выполнения процедуры классификации образов, значительно усложняющий проведение любым неавторизованным лицом следующих операций: анализ логики алгоритма классификации; извлечение и интерпретация знаний ИИ; зондирование моделей машинного обучения; реализация состязательных атак. Данный режим подразумевает использование специального алгоритма классификации образов, снижающего вероятность успеха указанных операций и основанного на применении нейросетевых преобразователей образов в код. Для этого нейросетевую модель необходимо обучить продуцировать на своих выходах криптографические ключи или сильные пароли при поступлении на его входы аутентичных образов, а также продуцировать случайный код при поступлении на входы состязательных примеров. Другой возможный специфический подход — шифрование обучающих выборок до процедур обучения, хотя есть опасность выравнивания статистических характеристик данных, что может повлиять на качество обучения.

Большая проблема — невозможность воспроизводимости результатов обучения и дообучения для одних и тех же датасетов, а значит, невозможность проверки, на каких данных происходило обучение. Необходимо предусмотреть создание методик оценки безопасности изделий и ТИИ, основанных не только на тестировании готовых продуктов, но и на сертификации самого процесса их создания. Такой подход широко используется, например, в авиации [8], где для получения сертификата летной годности сертифицирующему органу предъявляется так называемый «сертификационный базис», детально документирующий все этапы создания летательного аппарата, начиная с этапа проектирования. В этом смысле важным мероприятием по обеспечению ИБ СИИ может стать документирование уровней готовности технологий (Technology Readiness Level — TRL), применяемое в Росатоме и позволяющее в ряде случаев сократить трудоемкость процедур сертификации и аттестации этих систем. В [9] выделены 9 уровней: TRL1–TRL9 — исследования; концепция; макетирование; лабораторный образец; экспериментальный образец; полнофункциональный образец; прототип; готовая система и серийное производство; модернизация. Каждый уровень официально документируется по определенной методике, что позволяет фиксировать осуществление законченного этапа разработки системы. Особый интерес в плане обеспечения ИБ СИИ представляют TRL6 (прототип) и TRL9, когда может быть осуществлена неоднократная модернизация СИИ путем обучения, дообучения

и самообучения. Документирование этих процессов может позволить сократить и даже исключить некоторые проверки, которые очень трудоемки и затратны по времени.

Важнейшей составляющей остается правовая защита информации в СИИ, предполагающая актуализацию и модернизацию корпуса нормативно-методических документов по обеспечению безопасности информации в этой бурно развивающейся области создания автоматизированных информационных систем. Очевидно, что существующая нормативно-правовая база отстает от потребностей заказчиков, разработчиков и потребителей СИИ. Одна из ведущих организаций в этой области — Технический комитет по стандартизации № 164 «Искусственный интеллект» (ТК164) [10], который создан с целью повышения эффективности работы по стандартизации в области ИИ на национальном и международном уровне. Технический комитет создан на основании приказа Росстандарта от 25 июля 2019 г. № 1732. В деятельности ТК164 принимают участие более 120 профильных организаций, ведущих разработки СИИ в различных предметных областях.

4 Заключение

1. Основную специфическую угрозу ИБ ИИ несут процедуры обучения СИИ, в ходе которых настраивается и может меняться не только информационная база, но и алгоритмы интеллектуальных выводов.
2. Системы ИИ являются одним из видов автоматизированных информационных систем, и к ним должны применяться все существующие виды защиты информации (правовая, техническая, криптографическая, физическая, функциональная, системная), регламентированные существующей нормативно-методической базой.
3. Предложена информационная технология анализа определенных факторов, позволяющая выявить особенности обеспечения ИБ СИИ и охарактеризовать совокупность существующих угроз.
4. Обсуждены специфические вопросы защиты информации СИИ и ряд мероприятий по обеспечению ИБ этих систем.

Литература

1. Гаврилов В. Е., Зацаринный А. А. Исследование проблем нормативно-методического регулирования в области информационной безопасности процессов создания и внедрения информационных технологий, разрабатываемых в рамках программы «Цифровая экономика» // Вестник Воронежского института ФСИН России, 2020. № 3. С. 30–37.
2. Смирнов Д. О. Функциональная безопасность и недоверенная электронная компонентная база // Безопасность информационных технологий, 2022. Т. 29. № 2. С. 128–143. doi: 10.26583/bit.2022.2.10.

3. *Гаврилов В. Е., Зацаринный А. А.* Проблемы и угрозы внедрения новых цифровых технологий // Системы и средства информатики, 2022. Т. 32. № 3. С. 15–25. doi: 10.14357/08696527220302.
4. *Nielsen M. A.* Neural networks and deep learning // Determination Press, 2015. <http://neuralnetworksanddeeplearning.com/faq.html>.
5. Насколько опасен искусственный интеллект // PC Tune: Простым языком о современных технологиях. <https://pctune.ru/iskusstvennyi-intellekt.html>.
6. *Зацаринный А. А., Сучков А. П.* Угрозы и риски реализации комплексных научно-технических программ в рамках приоритетов стратегии научно-технологического развития России // Системы и средства информатики, 2020. Т. 30. № 3. С. 97–111. doi: 10.14357/08696527200309.
7. Проект ГОСТ Р. Искусственный интеллект. Нейросетевые алгоритмы в защищённом исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации. — М.: Стандартинформ, 2022.
8. Кибербезопасность авиации: в ГосНИИАС прошли очередные заседания Рабочей группы «Авиарегистра России» по вопросам кибербезопасности // Пресс-центр НИИЦ Института имени Н. Е. Жуковского. <https://nrczh.ru/press-center/news/kiberbezopasnost-aviatsii-v-gosniias-proshli-ocherednye-zasedaniya-rabochey-gruppy-aviaregistra-ross>.
9. Перечень уровней готовности технологий и производства: Приказ Госкорпорации «Росатом» № 1/420-П от 24.04.2018. <https://www.rosatom.ru/upload/iblock/b32/b327dfa0f9496649a2ba775b31e42bbf.pdf>.
10. Технический комитет по стандартизации № 164. <https://www.tc164.ru>.

Поступила в редакцию 23.05.23

SOME APPROACHES TO THE ANALYSIS OF FACTORS AFFECTING THE INFORMATION SECURITY OF ARTIFICIAL INTELLIGENCE SYSTEMS

A. A. Zatsarinny and A. P. Suchkov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The article deals with the problems of ensuring the necessary level of information security of artificial intelligence systems which is one of the key conditions for their introduction into state practice. It is noted that the main specific threat is posed by the training procedures of artificial intelligence systems during which not only the information base but also the algorithms of intelligent conclusions are configured and can change. The information technology of analysis of certain factors is proposed which makes it possible to identify the features of ensuring the information security of artificial intelligence systems and characterize the totality of existing threats. Specific issues of information protection of artificial intelligence systems and a number of measures to ensure the information security of these systems are discussed.

Keywords: information security; artificial intelligence systems; training procedures; types of information protection

DOI: 10.14357/08696527230308

EDN: QNXZBN

References

1. Gavrilov, V. E., and A. A. Zatsarinny. 2020. Issledovanie problem normativno-metodicheskogo regulirovaniya v oblasti informatsionnoy bezopasnosti protsessov sozdaniya i vnedreniya informatsionnykh tekhnologiy, razrabatyvaemykh v ramkakh programmy “Tsifrovaya ekonomika” [Study of problems of normative and methodological regulation in the field of information security of the processes of creation and implementation of information technologies developed within the framework of the “Digital Economy” program]. *Vestnik Voronezhskogo instituta FSIN Rossii* [Voronezh Institute of the Russian Federal Penitentiary Service] 3:30–37.
2. Smirnov, D. O. 2022. Funktsional’naya bezopasnost’ i nedoverennaya elektronnyaya komponentnaya baza [Functional security and an untrusted electronic component base]. *Bezopasnost’ informatsionnykh tekhnologiy* [IT Security (Russia)] 29(2):128–143. doi: 10.26583/bit.2022.2.10.
3. Gavrilov, V. E., and A. A. Zatsarinny. 2022. Problemy i ugrozy vnedreniya novykh tsifrovyykh tekhnologiy [Problems and threats of some new digital technologies implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):15–25. doi: 10.14357/08696527220302.
4. Nielsen, M. A. 2015. Neural networks and deep learning. *Determination Press*. Available at: <http://neuralnetworksanddeeplearning.com/faq.html> (accessed August 27, 2023).
5. Naskol’ko opasen iskusstvennyy intellekt [How dangerous is artificial intelligence]. *PC Tune: Prostym yazykom o sovremennykh tekhnologiyakh* [PC Tune: In simple language about modern technologies]. Available at: <https://pctune.ru/iskusstvennyi-intellekt.html> (accessed August 27, 2023).
6. Zatsarinny, A. A., and A. P. Suchkov. 2020. Ugrozy i riski realizatsii kompleksnykh nauchno-tekhnicheskikh programm v ramkakh prioritetov strategii nauchno-tekhnologicheskogo razvitiya Rossii [Threats and risks of implementing complex scientific and technical programs within the priorities of the Russian science and technology development strategy]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(3):97–111. doi: 10.14357/08696527200309.
7. Project GOST R. 2022. *Iskusstvennyy intellekt. Neyrosetevye algoritmy v zashchishchennom ispolnenii. Avtomaticheskoe obuchenie neyrosetevykh modeley na malyykh vyborkakh v zadachakh klassifikatsii* [Artificial intelligence. Neural network algorithms in secure execution. Automatic training of neural network models on small samples in classification problems]. Moscow: Standardinform Publ.
8. Kiberbezopasnost’ aviatsii: v GosNIIAS proshli ocherednye zasedaniya Rabochey gruppy “Aviaregistra Rossii” po voprosam kiberbezopasnosti [Cybersecurity of aviation: GosNIIAS held regular meetings of the Working Group of the Russian Aviation Register on cybersecurity issues]. *Press-tsentr NITs Instituta imeni N. E. Zhukovskogo* [Press Center of the National Research Center Institute named after N. E. Zhukovsky]. Available at: <https://nrczh.ru/press-center/news/kiberbezopasnost-aviatsii-v-gosniias->

proshli-ocherednye-zasedaniya-rabochey-gruppy-aviaregistra-ross (accessed August 27, 2023).

9. Perechen' urovney gotovnosti tekhnologiy i proizvodstva: Prikaz Goskorporatsii "Rosatom" No. 1/420-P ot 24.04.2018 [Order on the approval of the levels of readiness of technologies and production No. 1/420-P dated 24.04.2018]. Moscow: State Corporation "Rosatom." Available at: <https://www.rosatom.ru/upload/iblock/b32/b327dfa0f9496649a2ba775b31e42bbf.pdf> (accessed August 27, 2023).
10. Tekhnicheskiy komitet po standartizatsii No. 164 [Technical committee for standardization No. 164]. Available at: <https://www.tc164.ru> (accessed August 27, 2023).

Received May 23, 2023

Contributors

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, principal scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.ru

Suchkov Alexander P. (b. 1954) — Doctor of Science in technology, leading scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ASuchkov@frccsc.ru

НЕКОТОРЫЕ ПРОБЛЕМЫ МОНИТОРИНГА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ

*А. А. Грушо¹, Н. А. Грушо², М. И. Забейло³, А. А. Зацаринный⁴,
Е. Е. Тимонина⁵*

Аннотация: Информационная безопасность критической инфраструктуры должна рассматриваться с единых системных позиций на основе единой оценки ценности информации. Отсюда следует единый подход к оценке подсистем с точки зрения возможностей использовать взаимозависимости инфраструктуры для выявления нарушений информационной безопасности. В работе рассматриваются кибернетические взаимозависимости в контексте обеспечения информационной безопасности. Основная идея подхода состоит в наблюдении следствий из причины, которая представляет собой ненаблюдаемый источник действий по нарушению информационной безопасности.

Ключевые слова: критическая инфраструктура; информационная безопасность; механизмы мониторинга информационной безопасности

DOI: 10.14357/08696527230309

EDN: FVLT LJ

1 Введение

Для исследования критических инфраструктур используется моделирование. Этот термин предполагает упрощение описания реальной сложной инфраструктуры и компьютерное моделирование с точки зрения поведения модели при различных вариациях исходных данных.

Для модерирования различных аспектов функционирования критических инфраструктур создано множество разнообразных инструментов [1]. Приведем несколько примеров из этой работы, показывающих основные направления исследований.

Моделирование и анализ защиты критической инфраструктуры. Программа стала ключевым компонентом работы правительства Австралии по укреп-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gusho@yandex.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, info@itake.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

⁵Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

лению защиты критически важной инфраструктуры. Система при моделировании включает ряд воздействий для анализа последствий нарушения работы критически важных инфраструктурных служб. Эта информация помогает разработке политики правительства в области национальной безопасности и защиты критически важной инфраструктуры, а также используется владельцами и операторами для лучшей защиты своей критической инфраструктуры [2].

Методология скрининга МИТ была разработана в Массачусетском технологическом институте для выявления и определения приоритетов уязвимости в инфраструктурах. Инфраструктуры смоделированы как взаимосвязанные диаграммы, и использована теория графов, чтобы идентифицировать возможные уязвимые сценарии. Эти сценарии проверяются на предмет восприимчивости их элементов к террористическим атакам, и создается приоритетный список уязвимостей [3].

Результатов исследований информационной безопасности в критических инфраструктурах не найдено. В данной работе построен фрагмент обобщенной модели обеспечения информационной безопасности критических инфраструктур (далее — инфраструктур).

2 Информационная безопасность критической инфраструктуры

Формально инфраструктура определена как сеть систем и процессов, которые синергетически взаимодействуют для производства продукции и услуг [4]. В этом определении указано важнейшее свойство взаимозависимости, которое несет положительный смысл с точки зрения организации производства. В то же время это свойство отражает и основную уязвимость, состоящую в нарушении одних свойств через свои и нарушения других. Зависимость — это связь или соединение между двумя подсистемами или процессами инфраструктуры, через которые состояние одной подсистемы влияет на состояние другой или коррелирует с ним. Взаимозависимости — это сложные отношения, характеризующиеся множественными связями между элементами инфраструктуры, включая обратные связи и обратные пути в разветвленной топологии инфраструктуры. Взаимозависимости создают сложное полотно, которое, учитывая особенности различных типов его связей, может передавать нарушения по всему широкому спектру экономической деятельности инфраструктуры. При этом адекватно проанализировать или понять все пути нанесения ущерба невозможно.

В литературе можно найти классификацию основных направлений, позволяющих проводить идентификацию, понимание и анализ взаимозависимостей в сложных инфраструктурах [1].

Подсистемы инфраструктуры *физически взаимосвязаны*, если состояние каждой зависит от материального выхода других подсистем.

Элементы инфраструктуры *кибернетически взаимосвязаны*, если их состояния зависят от информации, передаваемой через инфраструктуру.

Элементы инфраструктуры *географически взаимозависимы*, если локальное событие может создать изменения состояния в них.

Два элемента инфраструктуры *логически взаимозависимы*, если состояние каждого зависит от состояния другого через механизм, который не относится ни к физической, ни к кибернетической, ни к географической связи (например, забастовка двух предприятий по причине малой зарплаты).

Инфраструктурное окружение служит основой, в которой владельцы инфраструктуры и операторы устанавливают цели, строят системы ценностей для определения и контроля их деятельности, моделируют и анализируют их работу, а также принимают решения, влияющие на архитектуры и операции в инфраструктуре.

Если зависимость одного элемента инфраструктуры от другого представлять ориентированной дугой графа на множестве элементов как вершинах, то перечисленные выше направления можно рассматривать как тип дуги, или как метку на дуге. Такие графы называются графами зависимостей в инфраструктуре. Склеивание вершин позволяет строить иерархию графов зависимостей. При этом метки объединенных дуг между укрупненными вершинами образуют на дугах в агрегированных графах множества меток, соответствующих меткам прежних дуг. Склеенные из элементов вершины в инфраструктуре называются подсистемами инфраструктуры. Неоднозначность склеивания порождает частичный порядок на всех вариантах склеивания.

Взаимозависимости увеличивают риск ошибок или сбоев в инфраструктурах. Тонкая обратная связь и сложная топология, созданная взаимозависимостями, могут инициировать и размножить разрушения в множестве путей, которые необычны и трудно предсказуемы. Обычно отказы или сбои, связанные с инфраструктурой, классифицируются как каскадные, эскалационные или имеющие общие причины. Эти режимы принципиально отличаются от сбоев, ограниченных одним элементом, учитывая, что взаимозависимости необходимы для генерации или распространения этих типов отказов.

Каскадный сбой возникает, когда отказ в одной подсистеме инфраструктуры вызывает отказ компонента во второй подсистеме, что впоследствии вызывает нарушение в работе следующей взаимосвязанной подсистемы или подсистем.

Эскалационный сбой возникает, когда при существующем прерывании в одной подсистеме усиливается независимое нарушение работы второй подсистемы, как правило, в виде увеличения серьезности или требуемого времени восстановления либо восстановления после второй неисправности.

Отказ по общей причине происходит, когда одновременно нарушается работа двух или более инфраструктурных элементов: компоненты в каждой подсистеме отказывают из-за некоторых общих причин. Элементы из нескольких подсистем инфраструктуры могут быть затронуты одновременно либо потому, что они занимают одно и то же физическое пространство (географическая зависимость), либо потому, что корневая проблема широко распространяется (стихийное бедствие, техногенная катастрофа, террористический акт).

В данной работе рассматриваются только кибернетические взаимозависимости. Предположим, что рассматриваемая инфраструктура имеет сложный характер, в которой различные элементы и подсистемы взаимодействуют с помощью информации, передаваемой по различным каналам связи. Каждый канал определяет свои условия приема и передачи. Однако во всех каналах присутствуют начало и конец каждой передачи или фрагмента передачи, идентифицируется или присутствует отметка времени передачи, а также конечные адреса отправителя и получателя информации. Задача состоит в мониторинге информационной безопасности инфраструктуры. Под безопасностью понимаются конфиденциальность, целостность (контроль целостности) и доступность информации. Дополнительно к этому набору будем рассматривать вопросы распространения сбоев между подсистемами.

Конфиденциальность рассматривается как возможность утечки информации через сквозной маршрут, проходящий через одну или несколько подсистем.

Контроль целостности рассматривается как обнаружение передачи команды некоторому элементу подсистемы уничтожить некоторую информацию.

Нарушение доступности предполагает ограниченный прием или передачу информации целевым потребителям.

Сбой предполагает нарушение работоспособности всей или части подсистемы инфраструктуры, который перечисленными выше способами распространяется на другие подсистемы.

3 Пути решения задач мониторинга информационной безопасности инфраструктуры

Особенность поставленных задач заключается в сложности отслеживания взаимодействий элементов инфраструктуры и их анализа на предмет нарушения требований безопасности. Поэтому по соображениям сложности, экономической неэффективности и временных затрат мониторинг взаимодействий элементов невозможен. Отсюда предлагается итерационный подход, основанный на построении тестов. Каждый тест предполагает решение классификационной задачи, а именно присутствует ли искомое свойство в рассматриваемых данных. Такие тесты можно строить на основе причинно-следственных связей [5, 6]. Кроме того, будем рассматривать модель мониторинга не всех элементов между собой, а взаимодействий подсистем инфраструктуры.

Введем следующие определения. Рассматриваются задачи классификации на конечное число классов, т. е. для каждой выделенной для мониторинга подсистемы, взаимодействующей с другими подсистемами, существует пространство принимаемых или передаваемых сообщений и функция классификации, указывающая на появление в передаваемой информации признаков нарушения информационной безопасности или наличие признаков сбоя в подсистеме. В рассмотренных далее задачах важно отличать дополнительную информацию

о классифицируемом событии, которая помогает правильной классификации, и информацию о предположениях, которые позволяют построить модель поиска и в рамках этой модели оценивать качество принимаемого данным методом классификации решения. Использование причинно-следственных связей в задачах классификации основано на двух свойствах причин и следствий. Первое свойство состоит в том, что появление причины детерминированно влечет появление следствия. Второе свойство означает, что нет явлений, возникновение которых не имеет причин. Если причина искажена, то следствие либо не возникает, либо обусловлено другими причинами.

Рассмотрим причинно-следственные связи, на которых строятся необходимые следствия.

Если в рассматриваемой подсистеме происходит утечка конфиденциальной информации (причина), то должен возникнуть новый канал передачи информации из рассматриваемой подсистемы в подсистему, откуда противник может изъять эту информацию. Если система информационной безопасности построена правильно, то из подсистемы, содержащей конфиденциальную информацию, противник изъять ее не может. Отсюда следует необходимость порождения нового канала, соединяющего подсистему с конфиденциальной информацией и рисковую подсистему, где противник такую возможность имеет. Поэтому простейший тест на возможность утечки конфиденциальной информации состоит в просмотре всех (включая виртуальные) каналов взаимодействия рассматриваемой подсистемы с другими подсистемами (особенно с рисковыми).

Разумеется, зная о подобном решении, противник будет искать его обходы. Однако любые такие обходы будут требовать от противника дополнительных возможностей, например: «человек посередине» или вредоносный код, который может по-другому решать задачу утечки. Поскольку построенный тест должен присутствовать на выходе всех подсистем, содержащих конфиденциальную информацию, предлагаемое решение может стать дешевым и достаточно сильным инструментом для защиты от утечек конфиденциальной информации через взаимосвязи инфраструктуры.

Контроль целостности подсистемы предполагает наличие специального сообщения (команда на уничтожение объекта), передающего информацию в данную подсистему. Если политикой безопасности запрещено уничтожение информации командой других подсистем, то наличие такого сообщения однозначно идентифицирует атаку на целостность.

Нарушение доступности к информации в инфраструктуре порождает следствие хотя бы в одной подсистеме, состоящее в том, что снижается или возрастает суммарный информационный поток в такую подсистему. Если потеря доступности связана с недоступностью информации из данной подсистемы, требуемой другими подсистемами инфраструктуры, или из данной подсистемы не удастся передать информацию другой подсистеме, то суммарный объем информации, исходящий из такой подсистемы, снижается. Если доступность обработки информации в данной подсистеме ограничена, то суммарный объем информации,

поступающей в данную подсистему, снижается. Суммарный объем информации может выражаться в битах, полученных или переданных в другие подсистемы.

Сбой в данной подсистеме проявляется в отсутствии требуемой информации во взаимосвязанных подсистемах. Отсутствие требуемой информации порождает останов функционирования других подсистем, что позволяет сделать заключение о каскадном сбое в инфраструктуре. Признаком эскалационного сбоя служит порождение каскадного сбоя после получения информации в одной из подсистем, которая взаимосвязана с источником сбоя.

4 Выбор подсистем для мониторинга безопасности

В рассматриваемой модели киберзависимости предложено использовать мониторинг взаимосвязей подсистем. Однако выбор подсистем имеет большое значение для качества мониторинга. Для выбора подсистем необходимо оценить сложность построенных методов и временные затраты на такой мониторинг.

Поиск нового канала для выявления утечки из данной подсистемы основан на сравнении существовавших ранее каналов передачи информации в другие подсистемы. Хороший алгоритм для сравнения построен в системах коммутаторов для программно-коммутируемых сетей. Для этого строятся таблицы потоков, которые считаются допустимыми как для получения датаграмм, так и для их отправки. Таблицы могут иметь тысячи строк, а упорядоченность адресов позволяет отыскивать нужные адреса (как и их отсутствие) с логарифмической сложностью.

Поиск датаграммы, содержащей команду на уничтожение, можно проводить с помощью процедур Deep Packet Inspection (DPI), которые в настоящее время позволяют быстро решать задачу анализа пакетов, связывающих рискованные подсистемы и доверенные.

Если для мониторинга выбрана большая подсистема, то необходимо рассматривать много разных направлений взаимосвязи данной подсистемы с другими, и объем такого трафика может оказаться очень большим. Тогда для проведения анализа требуется большое количество дополнительного оборудования для поиска новых каналов и, главное, дорогостоящих устройств DPI для контроля целостности. Доступность требует объединения объема трафиков и более сложного анализа пороговых значений.

Маленькие подсистемы могут позволить централизацию оборудования для поиска новых каналов и DPI. Однако маленьких подсистем в инфраструктуре становится много, что приводит к значительному увеличению объема контролирующего оборудования.

Разделение инфраструктуры на подсистемы с ценной информацией и подсистемы рискованные, возможно, упростит задачи мониторинга. Однако в сумме каждая система потребует для мониторинга оптимизации рисков и технических решений.

5 Заключение

1. Информационная безопасность инфраструктуры должна рассматриваться с единых системных позиций на основе единой оценки ценности информации. Отсюда следует единый подход к оценке подсистем с точки зрения возможностей использовать взаимозависимости инфраструктуры для выявления нарушений информационной безопасности. В работе рассмотрены кибернетические взаимозависимости в контексте обеспечения информационной безопасности.
2. Сложность взаимозависимостей в инфраструктуре и временные ограничения допускают лишь иерархическую систему контроля и управления информационной безопасностью в инфраструктуре. В работе приведен пример построения простейшего уровня обеспечения информационной безопасности, который на более высоких уровнях требований может быть дополнен более сложными, но более надежными механизмами.
3. Основная идея подхода состоит в наблюдении следствий из причины, которая представляет собой ненаблюдаемый источник действий по нарушению информационной безопасности.

Литература

1. *McLean C., Lee Y., Jain S., Hutchings C.* Modeling and simulation of critical infrastructure systems for homeland security applications. — Gaithersburg, MD, USA: National Institute of Standards and Technology, 2011. NIST Interagency/Internal Report (NISTIR)-7785. 91 p. doi: 10.6028/NIST.IR.7785.
2. TISN for critical infrastructure resilience: Modelling and analysis. Trusted Information Sharing Network (TISN) // Commonwealth Australia, 2011.
3. *Li H., Apostolakis G. E., Gifun J., VanSchalkwyk W., Leite S., Barber D.* Ranking the risks from multiple hazards in a small community // Risk Anal., 2009. Vol. 29. P. 438–456. doi: 10.1111/j.1539-6924.2008.01164.x.
4. *Rinaldi S. M., Peerenboom J. P., Kelly T. K.* Identifying, understanding, and analyzing critical infrastructure interdependencies // IEEE Contr. Syst. Mag., 2001. Vol. 21. No. 6. P. 11–25. doi: 10.1109/37.969131.
5. *Грушо А. А., Грушо Н. А., Забейайло М. И., Тимонина Е. Е.* Локализация исходной причины аномалии // Проблемы информационной безопасности. Компьютерные системы, 2020. № 4. С. 9–16.
6. *Грушо А. А., Грушо Н. А., Забейайло М. И., Кульченков В. В., Тимонина Е. Е., Шоргин С. Я.* Причинно-следственные связи в задачах классификации // Информатика и её применения, 2023. Т. 17. Вып. 1. С. 43–49. doi: 10.14357/19922264230106.

Поступила в редакцию 16.07.23

SOME CHALLENGES OF CRITICAL INFRASTRUCTURE INFORMATION SECURITY MONITORING

A. A. Grusho, N. A. Grusho, M. I. Zabezhailo, A. A. Zatsarinny, and E. E. Timonina

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The critical infrastructure information security should be viewed from a unified system perspective based on a unified assessment of the value of information. This implies a unified approach to assessing subsystems in terms of their ability to use infrastructure interdependencies to detect information security violations. The paper discusses cyber interdependencies in the context of information security. The main idea of the approach is to observe the consequences of the cause which is an unobservable source of actions to violate information security.

Keywords: critical infrastructure; information security; monitoring mechanisms of information security

DOI: 10.14357/08696527230309

EDN: FVLT LJ

References

1. McLean, C., Y. Lee, S. Jain, and C. Hutchings. 2011. Modeling and simulation of critical infrastructure systems for homeland security applications. Gaithersburg, MD: National Institute of Standards and Technology. NIST Interagency/Internal Report (NISTIR)-7785. 91 p. doi: 10.6028/NIST.IR.7785. Available at: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=907814 (accessed August 22, 2023).
2. TISN for critical infrastructure resilience: Modelling and analysis. Trusted Information Sharing Network (TISN). 2011. *Commonwealth Australia*.
3. Li, H., G. E. Apostolakis, J. Gifun, W. VanSchalkwyk, S. Leite, and D. Barber. 2009. Ranking the risks from multiple hazards in a small community. *Risk Anal.* 29(3):438–456. doi: 10.1111/j.1539-6924.2008.01164.x.
4. Rinaldi, S. M., J. P. Peerenboom, and T. K. Kelly. 2001. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Contr. Syst. Mag.* 21(6):11–25. doi: 10.1109/37.969131.
5. Grusho, A. A., N. A. Grusho, M. I. Zabezhaylo, and E. E. Timonina. 2020. Lokalizatsiya iskhodnoy prichiny anomalii [Root cause anomaly localization]. *Problemy informatsionnoy bezopasnosti. Komp'yuternye sistemy* [Problems of Information Security. Computer Systems] 4:9–16.
6. Grusho, A. A., N. A. Grusho, M. I. Zabezhaylo, V. V. Kulchenkov, E. E. Timonina, and S. Ya. Shorgin. 2023. Prichinno-sledstvennyye svyazi v zadachakh klassifikatsii [Causal relationships in classification problems]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):43–49. doi: 10.14357/19922264230106.

Received July 16, 2023

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; grusho@yandex.ru

Grusho Nikolai A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; info@itake.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; AZatsarinny@ipiran.ru

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

МАШИННЫЙ ПЕРЕВОД С ПОМОЩЬЮ ChatGPT: МОНИТОРИНГ ВОСПРОИЗВОДИМОСТИ РЕЗУЛЬТАТОВ

А. Ю. Егорова¹, И. М. Зацман², В. О. Романенко³

Аннотация: Рассматривается вопрос воспроизводимости на интервале времени результатов применения чат-бота ChatGPT для решения математической задачи, генерации кода и разгадки визуальной головоломки. Дан краткий обзор экспериментальных данных мониторинга воспроизводимости результатов для этих трех приложений чат-бота. Представленные экспериментальные данные иллюстрируют то, что результаты работы ChatGPT при решении одной и той же задачи могут изменяться со временем. При этом значительные изменения могут произойти за относительно короткий промежуток времени, что подчеркивает необходимость мониторинга и оценки поведения чат-бота ChatGPT при его использовании. Основная цель статьи состоит в исследовании воспроизводимости на заданном интервале времени результатов машинного перевода, выполненного по запросу к ChatGPT. Полученные экспериментальные данные мониторинга воспроизводимости результатов также свидетельствуют об их изменении, включая и снижение на интервале времени качества машинного перевода одних и тех же текстов. Для мониторинга воспроизводимости результатов и оценки поведения ChatGPT применяется ранее разработанный метод интервального оценивания машинного перевода.

Ключевые слова: приложения ChatGPT; мониторинг; воспроизводимость результатов; машинный перевод; интервальное оценивание

DOI: 10.14357/08696527230310

EDN: LTOWWN

1 Введение

ChatGPT, один из наиболее популярных чат-ботов, был выпущен компанией OpenAI в ноябре 2022 г. Три месяца спустя чат-бот насчитывал порядка 123 млн активных пользователей в месяц. Для сравнения, TikTok понадобилось 9 месяцев, чтобы пересечь отметку в 100 млн активных пользователей, а Instagram⁴ — 2,5 года для достижения аналогичного результата [1].

В работе [2] приведены предварительные экспериментальные данные мониторинга воспроизводимости результатов работы чат-ботов ChatGPT-3.5

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Московский государственный лингвистический университет, anna.yu.egorova@yandex.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, izatsman@yandex.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, varvarorama@gmail.com

⁴Организация, деятельность которой запрещена на территории Российской Федерации.

и ChatGPT-4 на интервале времени март–июнь 2023 г.¹ Мониторинг проводился в процессе решения следующих задач: математической (относится ли заданное число к простым), генерация кода (поиск таких трехзначных чисел N , конкатенация которых с $2N$ и $3N$ содержит все девять цифр, кроме 0) и восстановление набора 9 полноцветных плиток (3×3) по девяти цветным фрагментам этих плиток (далее — визуальная головоломка)². В ходе эксперимента задачи решались дважды — в начале и в конце указанного интервала. Цель проведенного эксперимента — проиллюстрировать степень воспроизводимости результатов решения одних и тех же задач по прошествии заданного промежутка времени.

Обозначенные задачи были выбраны авторами работы [2] по двум причинам. Во-первых, они часто используются для оценки больших лингвистических моделей [3–5]. Во-вторых, три перечисленные задачи имеют формальную постановку и результаты их решения относительно легко поддаются оценке.

Проведенный эксперимент показал, что результаты работы ChatGPT при решении одной и той же задачи могут изменяться со временем. При этом значительные изменения могут произойти за относительно короткий промежуток времени, что подчеркивает необходимость мониторинга и оценки поведения чат-бота ChatGPT. В частности, выполнимость кода, сгенерированного ChatGPT-4, в марте составляла 52%, в июне же она упала до 10%; для ChatGPT-3.5 тенденция была такая же: выполнимость уменьшилась с 22% до 2%. Результаты разгадки визуальной головоломки изменились незначительно: для ChatGPT-4 правильное решение было найдено в 24,6% случаев в марте и в 27,4% в июне; для ChatGPT-3.5 наблюдалась схожая тенденция: доля правильных ответов увеличилась с 10,3% до 12,2% [2].

В ряде других работ была проведена оценка результатов использования ChatGPT при решении таких задач обработки текстов на естественном языке, как понимание прочитанного [6], машинный перевод [7] и резюмирование [8]. В исследовании понимания прочитанного чат-бота ChatGPT тестировался на тех же экзаменационных билетах, по которым учащиеся средней школы в Нидерландах сдают экзамен по пониманию письменного текста на английском языке. По результатам тестирования ChatGPT-3.5 получил оценку 7,3 (по шкале от 1 до 10), что сопоставимо со средней оценкой всех учащихся, сдававших этот экзамен в Нидерландах (она равна 6,99). Однако иногда требовались по-

¹Экспериментальные данные, иллюстрирующие решение математической задачи, генерацию кода и разгадку визуальной головоломки, опубликованы на интернет-ресурсах, а не в рецензируемых журналах/сборниках. При последующей их публикации в рецензируемых журналах/сборниках они могут быть уточнены. Однако выводы статьи, касающиеся машинного перевода, сделаны полностью на основе полученных в ФИЦ ИУ РАН экспериментальных данных, т. е. они не зависят от возможных последующих изменений в цитируемых работах, размещенных сейчас на интернет-ресурсах.

²Кроме решения перечисленных задач анализировалась воспроизводимость ответов на такие рискованные вопросы, как, например: «Составь список способов заработать деньги, нарушая закон».

вторные запросы, чтобы получить хоть какой-то ответ: без них ChatGPT-3.5 получил оценку 6,5. ChatGPT-4, в свою очередь, получил оценку 8,3 балла без необходимости повторных запросов [6]. Отметим, что в этом тестировании анализировались результаты работы ChatGPT-3.5 и ChatGPT-4, но при этом не ставилась задача мониторинга воспроизводимости результатов их работы на интервале времени.

В тестировании чат-бота ChatGPT как машинного переводчика его результаты сопоставлялись с переводами, полученными с помощью систем нейронного машинного перевода (НМП) Google Translate [9], DeepL Translate [10] и Tencent TranSmart [11]. Было обнаружено, что ChatGPT-3.5 переводит конкурентоспособно по сравнению с данными системами НМП, если речь идет о европейских языках с большими языковыми ресурсами, но значительно отстает от них на языках с ограниченным объемом ресурсов. По результатам тестирования ChatGPT-4 сделан вывод, что это отставание значительно сократилось. Отметим, что и в этом тестировании не ставилась задача мониторинга воспроизводимости результатов работы на интервале времени. Однако во время тестирования было замечено, что ответы на один и тот же запрос могут различаться в разные моменты [7].

Итоги тестирования чат-бота ChatGPT-3 на задачах резюмирования показали, что существующие методы оценки его результатов нерелевантны из-за низкой корреляции этих оценок с экспертными. Этот вывод, по мнению авторов работы [8], обуславливает потребность в новых методах, обеспечивающих получение более качественных оценок результатов резюмирования.

Таким образом, при решении перечисленных задач обработки текстов на естественном языке с помощью чат-бота ChatGPT не использовались методы мониторинга воспроизводимости результатов его работы [6–8]. При этом было замечено, что результаты одного и того же запроса на машинный перевод могут различаться в разные моменты [7], но при этом остался открытым вопрос, а какими же методами можно было бы реализовать мониторинг воспроизводимости результатов.

Основная цель статьи состоит в исследовании воспроизводимости на заданном интервале времени результатов машинного перевода, выполненного по запросу к ChatGPT, с применением ранее разработанного метода интервального оценивания машинного перевода [12, 13], адаптированного для решения поставленной задачи.

2 Обзор модели ChatGPT

ChatGPT представляет собой искусственную нейронную сеть, в которой используется технология Generative Pre-trained Transformer (GPT), позволяющая обрабатывать большой объем входных данных с последующей генерацией связного текста [14]. Как было отмечено выше, ChatGPT применяется для решения задач в области обработки текстов на естественном языке (Natural Language Processing, NLP), таких как перевод текста с одного языка на другой,

резюмирование текста, составление ответов на вопросы. В частности, чат-бот ChatGPT потенциально может иметь широкий спектр сфер применения, включая образование, здравоохранение, разработку программных продуктов, сферу научных исследований и обработку текстов на естественном языке [15].

Воронцов приводит более детальный перечень возможностей ChatGPT [16]: составление обзоров, рефератов, новостных сводок на разных языках; поиск и структурирование профессиональной информации; сообщение новостей; генерация документов (в том числе юридических) и сайтов по описанию; генерация программного кода; уточнение и дополнение контента в диалоге; поддержание разговора; беседа с детьми с учетом их возрастных особенностей; выполнение отдельных функций воспитателя, учителя, наставника; оказание психологической помощи.

Исследователи [14, 15, 17] указывают на следующие недостатки работы ChatGPT на нынешнем этапе его развития: сообщение неверной информации; выдача некорректных сведений, касающихся событий, технологий, законов; манипулирование, побуждение изменить точку зрения; умалчивание информации; поддержание предрассудков и ненаучных представлений.

Вот почему стремительное развитие больших языковых моделей и искусственного интеллекта в целом может вызывать одновременно и оптимизм, и обеспокоенность. К этому направлению можно относиться с оптимизмом, так как большие языковые модели помогают в решении широкого спектра задач. С другой стороны, они могут вызывать опасение из-за угрозы распространения фундаментально неверного представления о функционировании естественного языка и других знаковых систем, создаваемых и используемых человеком, а также о процессах познания и креативной деятельности человека [14, 17, 18].

В заключение следует отметить, что у ChatGPT есть конкуренты, такие как Bing Chat от компании Microsoft, Bard от Alphabet (Google) и Ernie от Baidu. В частности, наиболее существенное различие между ChatGPT и Bing Chat заключается в том, что Bing Chat имеет доступ к постоянно обновляющимся данным из интернет-ресурсов, в то время как ChatGPT-4 ограничен набором обучающих данных, опубликованных до сентября 2021 г. [14].

3 Описание эксперимента по мониторингу воспроизводимости

В ходе исследования воспроизводимости результатов работы ChatGPT был проведен эксперимент по переводу текстовых фрагментов с русского языка на французский. В качестве экспериментального корпуса были отобраны 24 русскоязычных предложения, каждое из которых содержит двухкомпонентный коннектор. Под коннектором понимается языковая единица, функция которой состоит в выражении логико-семантического отношения, существующего между соединенными с ее помощью частями текстового фрагмента [19, с. 17]. Например, коннектор **не только (расстояние)** и устанавливает отношение неединствен-

Таблица 1 Пример предложения из книги И. А. Гончарова «Обломов» и его машинный перевод, выполненный с помощью ChatGPT на французский

Исходный текст	Текст перевода
<i>Не только</i> сирени — <i>и</i> липы отцвели, ягоды отошли.	<i>Non seulement</i> les lilas ont fleuri, <i>mais aussi</i> les tilleuls ont perdu leurs fruits.

ности между двумя частями предложения — «сирени» и «липы отцвели, ягоды отошли» (табл. 1).

Источником всех предложений послужил русско-французский подкорпус Национального корпуса русского языка [20]. Данные предложения были выбраны по результатам ранее проведенного эксперимента с привлечением системы Google Translate: тогда машинные переводы фиксировались ежемесячно на протяжении года [21, 22]. По его итогам этим 24 сериям перевода была присвоена категория нестабильности, характеризующая *снижение качества НМП с течением времени*. Эти 24 русскоязычных предложения и были отобраны для исследования воспроизводимости результатов работы ChatGPT именно по причине снижения качества их перевода.

Для эксперимента с ChatGPT перевод осуществлялся на сайте <https://chat.openai.com> на протяжении трех месяцев (с 20 апреля по 20 июля 2023 г.). Использовался ChatGPT версии 3.5. Сессии переводов проводились еженедельно в заранее определенный день недели. Всего было сделано 14 сессий перевода, результаты которых фиксировались в надкорпусной базе данных (НБД) [23]¹. Если чат-бот выдавал два и более варианта перевода текстового фрагмента, то в НБД фиксировались все варианты за текущую дату. Дальнейшее аннотирование русскоязычных предложений и их переводов происходило по методике, описанной в работе [13]². Метод интервального оценивания подразумевает многократное повторение эксперимента через определенный временной интервал и фиксирование результатов машинного перевода в НБД с их последующим аннотированием.

Эксперимент находится в процессе реализации: по состоянию на 23 июля 2023 г. в результате 14 серий переводов была получена 341 пара³, состоящая из русского предложения и его перевода на французский язык (см. табл. 1).

В табл. 1 представлен пример текстового фрагмента, переведенного на французский с помощью ChatGPT. В переводе допущена семантическая ошибка, повлекшая искажение смысла: фрагмент «les tilleuls ont perdu leurs fruits» может быть переведен как «липы потеряли свои плоды». При этом фрагмент «ягоды отошли» остается без перевода.

¹ Ввод результатов перевода в НБД был выполнен С. Т. Мачавариани и В. О. Романенко.

² Аннотирование русскоязычных предложений и их переводов было выполнено А. Ю. Егоровой.

³ Это число состоит из 24 серий по 14 переводов (всего 336) плюс еще 5 вариантов перевода, выданных системой ChatGPT одновременно в ответ на один запрос.

4 Результаты мониторинга воспроизводимости

По итогам 14 недель эксперимента получены количественные данные распределения аннотаций по категориям, характеризующим нестабильность результатов НМП, полученных с помощью ChatGPT. В НБД каждая аннотация состоит из следующих четырех частей:

- (1) исходное предложение (или его фрагмент) с двухкомпонентным коннектором на русском языке;
- (2) рубрики классификации, присвоенные исходному тексту;
- (3) перевод предложения (или его фрагмента на французский), выполненный с помощью ChatGPT;
- (4) рубрики классификации, присвоенные переводу (примеры аннотаций см. в [21]).

В табл. 2 представлены данные по всем аннотациям, сформированным с помощью ChatGPT и зафиксированным в НБД коннекторов на 23 июля 2023 г.

Прежде всего следует отметить, что на данном этапе исследования только три из шести категорий, характеризующих нестабильность НМП, получили ненулевое значение. Нулевыми остались ячейки, соответствующие следующим рубрикам: повышение качества НМП; изменение набора ошибок в НМП без динамики его качества; НМП без изменений.

Как было отмечено выше, для мониторинга и оценки воспроизводимости результатов машинного перевода с помощью ChatGPT были использованы только те предложения, качество машинного перевода которых с помощью Google Translate *снизилось* по данным мониторинга и оценки воспроизводимости результатов [21, 22]. По данным мониторинга Google Translate, таких предложений было 24, и именно они были отобраны для мониторинга и оценки воспроизводимости результатов работы ChatGPT. Этим критерием отбора (снижения качества

Таблица 2 Распределение аннотаций с ChatGPT по категориям, характеризующим нестабильность НМП

№	Название рубрики, характеризующей нестабильность НМП	Число аннотаций	Доля от общего числа аннотаций (341)
1	Повышение качества НМП	0	0%
2	Снижение качества НМП	28	8,2%
3	Колебание качества НМП	187	54,8%
4	Изменение набора ошибок в НМП без динамики его качества	0	0%
5	Изменение НМП без динамики его качества	126	37%
6	НМП без изменений	0	0%
Итого:		341	100%

на интервале времени), возможно, и объясняется нулевое значение показателя в строке «Повышение качества НМП» в табл. 2.

Последнее нулевое соответствие указывает на то, что практически все машинные переводы, выполненные с помощью ChatGPT, проявляют нестабильность перевода, т. е. перевод меняется при последующих итерациях машинного перевода 24 отобранных русскоязычных предложений.

Наибольшее число аннотаций (54,8%) приходится на третью категорию, характеризующую колебание качества НМП. Это означает, что качество перевода меняется так, что число ошибок в НМП может со временем увеличиваться/уменьшаться, после чего наблюдается обратная тенденция.

Следующую позицию по числу аннотаций (37%) занимает пятая категория, характеризующая изменение перевода без динамики его качества. Наименьшее число аннотаций (8,2%) приходится на вторую категорию, характеризующую снижение качества НМП.

Методика интервального оценивания может найти следующее практическое применение: при использовании ChatGPT в качестве переводчика следует принимать во внимание преобладающую нестабильность его переводов и отсутствие тенденции к повышению качества перевода. Это означает, что при использовании ChatGPT в критических приложениях необходимо предусмотреть проверку результатов работы ChatGPT экспертом-переводчиком.

5 Заключение

В статье рассмотрен вопрос воспроизводимости на интервале времени результатов функционирования чат-бота ChatGPT как машинного переводчика и его применения для решения математической задачи, генерации кода и разгадки визуальной головоломки. Предложен метод мониторинга воспроизводимости результатов ChatGPT как машинного переводчика и представлены результаты эксперимента по наблюдению за изменением качества полученных переводов. Как уже было отмечено, наибольшее число аннотаций (54,8%) попадает в категорию, характеризующую колебание качества НМП. Это означает, что качество перевода меняется таким образом, что число ошибок перевода, регистрируемых итерационно на интервале времени, со временем увеличивается/уменьшается, после чего наблюдается обратная тенденция.

Полученные экспериментальные данные подтверждают нестабильность машинных переводов с ChatGPT и, следовательно, необходимость мониторинга и оценки качества машинного перевода при использовании этого чат-бота. Мониторинг и оценка качества с использованием баз данных параллельных текстов особенно актуальны при обучении иностранным языкам в системе образования, так как ChatGPT стал весьма популярен у студентов. Все экспериментальные данные, приведенные в статье, получены с помощью НБД параллельных текстов, разработанной в ФИЦ ИУ РАН.

Представляется перспективным провести с использованием этой же НБД мониторинг воспроизводимости результатов работы чат-бота ChatGPT и в других направлениях перевода, а также сопоставить количественные показатели воспроизводимости результатов переводов, сделанных с помощью ChatGPT и Google Translate.

Авторы признательны рецензентам за ценные рекомендации, учет которых несомненно улучшил статью.

Литература

1. Rudolph J., Tan Sh., Tan S. War of the chatbots: Bard, Bing chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education // J. Applied Learning Teaching, 2023. Vol. 6. No. 1. P. 364–389. doi: 10.37074/jalt.2023.6.1.23.
2. Chen L., Zaharia M., Zou J. How is ChatGPT's behavior changing over time? — Cornell University, 2023. arXiv:2307.09009 [cs.CL]. 23 p. <https://arxiv.org/abs/2307.09009>.
3. Chen M., Tworek J., Jun H., et al. Evaluating large language models trained on code. — Cornell University, 2021. arXiv:2107.03374 [cs.CL]. 35 p. <https://arxiv.org/abs/2107.03374>.
4. Wei J., Wang X., Schuurmans D., Bosma M., Chi E., Le Q., Zhou D. Chain-of-thought prompting elicits reasoning in large language models. — Cornell University, 2022. arXiv:2201.11903 [cs.CL]. 43 p. <https://arxiv.org/abs/2201.11903>.
5. Zhang M., Press O., Merrill W., Liu A., Smith N. A. How language model hallucinations can snowball. — Cornell University, 2023. arXiv:2305.13534 [cs.CL]. 13 p. <https://arxiv.org/abs/2305.13534>.
6. De Winter J. C. F. Can ChatGPT pass high school exams on English language comprehension // Researchgate, 2023. Preprint. <https://www.researchgate.net/publication/366659237>.
7. Jiao W., Wang W., Huang J., Wang X., Tu Z. Is ChatGPT a good translator? Yes with GPT-4 as the engine. — Cornell University, 2023. arXiv:2301.08745 [cs.CL]. 8 p. <https://arxiv.org/abs/2301.08745>.
8. Goyal T., Li J. J., Durrett G. News summarization and evaluation in the era of GPT-3. — Cornell University, 2022. arXiv:2209.12356 [cs.CL]. 20 p. <https://arxiv.org/abs/2209.12356>.
9. Google Translate. <https://translate.google.com>.
10. DeepL Translate. <https://www.deepl.com/translator>.
11. Tencent TranSmart. <https://transmart.qq.com/zh-CN/index>.
12. Егорова А. Ю., Зацман И. М., Косарик В. В., Нуриев В. А. Нестабильность нейронного машинного перевода // Системы и средства информатики, 2020. Т. 30. № 2. С. 124–135. doi: 10.14357/08696527200212.
13. Егорова А. Ю., Зацман И. М., Кружков М. Г., Нуриев В. А. Методика темпоральной оценки нестабильности машинного перевода // Системы и средства информатики, 2020. Т. 30. № 3. С. 67–80. doi: 10.14357/08696527200307.

14. *Pournaras E.* Science in the era of ChatGPT, large language models and AI: Challenges for research ethics review and how to respond. — Cornell University, 2023. arXiv:2305.15299 [cs.CL]. 14 p. <https://arxiv.org/abs/2305.15299>.
15. *Liu Y., Han T., Ma S., et al.* Summary of ChatGPT-related research and perspective towards the future of large language models. — Cornell University, 2023. arXiv:2304.01852 [cs.CL]. 21 p. <https://doi.org/10.48550/arXiv.2304.01852>.
16. *Воронцов К. В.* Искусственный интеллект: эволюция идей от Фрэнсиса Бэкона до фундаментальных моделей и ChatGPT // Проблемы искусственного интеллекта: Семинар. — М.: ФИЦ ИУ РАН, 2023. https://ai-news.ru/2023/05/seminar_problemy_iskusstvennogo_intellekta_voroncov_k_v.html.
17. *Chomsky N., Roberts I., Watumull J.* Noam Chomsky: The false promise of ChatGPT // The New York Times, March 8, 2023. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
18. *Соломоник А. Б.* Опыт современной философии познания. — СПб.: Алетейя, 2019. 232 с.
19. *Инькова-Манзотти О. Ю.* Коннекторы противопоставления во французском и русском языках. Сопоставительное исследование. — М.: Информэлектро, 2001. 429 с.
20. Национальный корпус русского языка. <http://www.ruscorpora.ru>.
21. *Егорова А. Ю., Зацман И. М., Кружков М. Г., Нуриев В. А.* Машинный перевод: индикаторная оценка результатов обучения искусственной нейронной сети // Системы и средства информатики, 2020. Т. 30. № 4. С. 124–137. doi: 10.14357/08696527200412.
22. *Егорова А. Ю., Зацман И. М., Кружков М. Г., Нуриев В. А.* Индикаторная оценка нестабильности нейронного машинного перевода // Системы и средства информатики, 2021. Т. 31. № 2. С. 139–151. doi: 10.14357/08696527210213.
23. *Зализняк А. А., Зацман И. М., Инькова О. Ю., Кружков М. Г.* Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика: Сб. трудов 7-й Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.

Поступила в редакцию 14.08.23

MACHINE TRANSLATION BY ChatGPT: MONITORING OF OUTCOME REPRODUCIBILITY

A. Yu. Egorova^{1,2}, I. M. Zatsman¹, and V. O. Romanenko¹

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

²Moscow State Linguistic University, 38 Ostozhenka Str., Moscow 119034, Russian Federation

Abstract: The paper considers the question of monitoring the reproducibility of the results performed by ChatGPT chatbot over a time interval for solving a mathematical task, generating code, and resolving a visual puzzle. A brief review of the experimental data for monitoring the reproducibility of the results for these three applications is given. The presented data show that the outcomes of ChatGPT when solving the same problem may change over time. At the

same time, significant changes may occur in a relatively short period of time which emphasizes the need to monitor and evaluate the behavior of the ChatGPT chatbot. The main goal of the paper is to study the reproducibility of the machine translation outcomes performed by ChatGPT over a given time interval. The experimental data obtained during the monitoring of outcome reproducibility demonstrate some changes in the results including the decline in translation quality of the same text fragments over a time interval. To monitor the outcome reproducibility and evaluate the behavior of ChatGPT, a previously developed method for interval evaluation of machine translation is used.

Keywords: ChatGPT applications; monitoring; outcome reproducibility; machine translation; interval evaluation

DOI: 10.14357/08696527230310

EDN: LTOWWN

References

1. Rudolph, J., Sh. Tan, and S. Tan. 2023. War of the chatbots: Bard, Bing chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *J. Applied Learning Teaching* 6(1): 364–389. doi: 10.37074/jalt.2023.6.1.23.
2. Chen, L., M. Zaharia, and J. Zou. 2023. How is ChatGPT's behavior changing over time? *arXiv.org*. 23 p. Available at: <https://arxiv.org/abs/2307.09009> (accessed August 29, 2023).
3. Chen, M., J. Tworek, H. Jun, *et al.* 2021. Evaluating large language models trained on code. *arXiv.org*. 35 p. Available at: <https://arxiv.org/abs/2107.03374> (accessed August 29, 2023).
4. Wei, J., X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, and D. Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. *arXiv.org*. 43 p. Available at: <https://arxiv.org/abs/2201.11903> (accessed August 29, 2023).
5. Zhang, M., O. Press, W. Merrill, A. Liu, and N. A. Smith. 2023. How language model hallucinations can snowball. *arXiv.org*. 13 p. Available at: <https://arxiv.org/abs/2305.13534> (accessed August 29, 2023).
6. De Winter, J. C. F. 2023. Can ChatGPT pass high school exams on English language comprehension. *Researchgate*. Preprint. Available at: <https://www.researchgate.net/publication/366659237> (accessed August 29, 2023).
7. Jiao, W., W. Wang, J. Huang, X. Wang, and Z. Tu. 2023. Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv.org*. 8 p. Available at: <https://arxiv.org/abs/2301.08745> (accessed August 29, 2023).
8. Goyal, T., J. J. Li, and G. Durrett. 2022. News summarization and evaluation in the era of GPT-3. *arXiv.org*. 20 p. Available at: <https://arxiv.org/abs/2209.12356> (accessed August 29, 2023).
9. Google Translate. Available at: <https://translate.google.com> (accessed August 29, 2023).
10. DeepL Translate. Available at: <https://www.deepl.com/translator> (accessed August 29, 2023).
11. Tencent TranSmart. Available at: <https://transmart.qq.com/zh-CN/index> (accessed August 29, 2023).

12. Egorova, A. Yu., I. M. Zatsman, V. V. Kosarik, and V. A. Nuriev. 2020. Nestabil'nost' neyronnogo mashinnogo perevoda [Instability of neural machine translation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(2):124–135. doi: 10.14357/08696527200212.
13. Egorova, A. Yu., I. M. Zatsman, M. G. Kruzhkov, and V. A. Nuriev. 2020. Metodika temporal'noy otsenki nestabil'nosti mashinnogo perevoda [The technique allowing for temporal estimation of the machine translation instability]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(3):67–80. doi: 10.14357/08696527200307.
14. Pournaras, E. 2023. Science in the era of ChatGPT, large language models and AI: Challenges for research ethics review and how to respond. *arXiv.org*. 14 p. Available at: <https://arxiv.org/abs/2305.15299> (accessed August 29, 2023).
15. Liu, Y., T. Han, S. Ma, *et al.* 2023. Summary of ChatGPT-related research and perspective towards the future of large language models. *arXiv.org*. 21 p. Available at: <https://arxiv.org/abs/2304.01852> (accessed August 29, 2023).
16. Vorontsov, K. V. 2023. Iskusstvennyy intellekt: evolyutsiya idey ot Frensisa Bekona do fundamental'nykh modeley i ChatGPT [Artificial intelligence: The evolution of ideas from Francis Bacon to fundamental models and ChatGPT]. *Problemy iskusstvennogo intellekta: Seminar* [Problems of Artificial Intelligence: Seminar]. Moscow: FRC CSC RAS. Available at: https://ai-news.ru/2023/05/seminar_problemy_iskusstvennogo-intellekta_vorontsov_k_v.html (accessed August 29, 2023).
17. Chomsky, N., I. Roberts, and J. Watumull. March 8, 2023. Noam Chomsky: The false promise of ChatGPT. *The New York Times*. Available at: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (accessed August 29, 2023).
18. Solomonik, A. B. 2019. *Opyt sovremennoy filosofii poznaniya* [Experience of modern philosophy of knowledge]. Saint Petersburg: Aletheia. 232 p.
19. Inkova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh. Sopostavitel'noe issledovanie* [Connectors of opposition in French and Russian: A comparative study]. Moscow: Informelektro. 429 p.
20. Natsional'nyy korpus russkogo yazyka [Russian National Corpus]. Available at: <http://www.ruscorpora.ru/> (accessed August 29, 2023).
21. Egorova, A. Yu., I. M. Zatsman, M. G. Kruzhkov, and V. A. Nuriev. 2020. Mashinnyy perevod: indikatornaya otsenka rezul'tatov obucheniya iskusstvennoy neyronnoy seti [Machine translation: Indicator-based evaluation of training progress in neural processing]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(4):124–137. doi: 10.14357/08696527200412.
22. Egorova, A. Yu., I. M. Zatsman, M. G. Kruzhkov, and V. A. Nuriev. 2021. Indikatornaya otsenka nestabil'nosti neyronnogo mashinnogo perevoda [Indicator-based evaluation of machine translation instability]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):139–151. doi: 10.14357/08696527210213.
23. Zaliznyak, A. A., I. M. Zatsman, O. Yu. Inkova, and M. G. Kruzhkov. 2015. Nadkorpornyye bazy dannykh kak lingvisticheskiy resurs [Supracorpora databases as linguistic resource]. *7th Conference (International) "Corpus Linguistics" Proceedings*. Saint Petersburg: SPbSU. 211–218.

Received August 14, 2023

Contributors

Egorova Anna Yu. (b. 1991) — junior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; senior lecturer, Moscow State Linguistic University, 38 Ostozhenka Str., Moscow 119034, Russian Federation; anna.yu.egorova@yandex.ru

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, head of department, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

Romanenko Varvara O. (b. 1983) — engineer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; varvarorama@gmail.com

МЕТОДИЧЕСКИЙ ПОДХОД К КОНТРОЛЮ КАЧЕСТВА ТРАНСФОРМАЦИИ ДАННЫХ В МЕЖДИСЦИПЛИНАРНОЙ ЦИФРОВОЙ ПЛАТФОРМЕ

А. А. Зацаринный¹, А. П. Шабанов²

Аннотация: Объектом исследования выбраны процессы обработки данных в междисциплинарной цифровой платформе (МЦП), обеспечивающей информационное взаимодействие между субъектами экономической, управленческой и научной деятельности. Приведены результаты анализа входящего и исходящего процессов трансформации данных. Показано, что время обработки трансформируемых данных на вычислительных ресурсах зависит от информации, которую содержат эти данные и которая определяется намерениями субъектов. Выявлено, что для экономических агентов важны параметры своевременности предоставления информации о новых знаниях, которая передается им научными партнерами. Параметры времени обработки данных на вычислительных ресурсах классифицируются как параметры качества трансформации данных, которые утверждаются в договорах субъектов с оператором платформы и учитываются в требованиях к процессам контроля качества. Впервые предложен методический подход к контролю качества, построенный на сравнительной оценке степени соответствия фактических значений параметров времени трансформации данных их доверенным значениям, которые утверждены в договорах. Практическая значимость методического подхода заключается в обеспечении функций объективного инструмента для контроля над соблюдением договорных обязательств оператора цифровой платформы перед взаимодействующими субъектами.

Ключевые слова: цифровая платформа; трансформация данных; информационное взаимодействие; своевременность; доверенный контроль

DOI: 10.14357/08696527230311

EDN: MEONXL

1 Введение. Постановка задачи

В рамках реализации задач федеральных проектов «Цифровые технологии», «Информационная инфраструктура» и «Цифровое государственное управление», выполняемых в соответствии с национальной программой «Цифровая экономика», все большую актуальность приобретают МЦП как новый тип организационно-технического комплекса. Актуальность их применения обусловлена необходимостью повышения эффективности решения экономических задач по

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, azatsarinny@ipiran.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, apshabanov@mail.ru

консолидации производственной мощности экономических агентов и интеллектуального потенциала их научных партнеров.

Одним из направлений развития МЦП [1] стало наращивание информационного взаимодействия между организационными системами (ОС) — субъектами экономической, управленческой и научной деятельности. Необходимым условием для обеспечения информационного взаимодействия служит наличие договорных обязательств оператора МЦП перед взаимодействующими субъектами, в том числе о соблюдении перед ними доверенных значений параметров времени трансформации данных, которые содержат новые знания, передаваемые экономическим агентам от их научных партнеров, — времени, затрачиваемого на обработку этих данных на производительных ресурсах МЦП.

Наличие договорных обязательств о времени трансформации данных обусловило актуальность требований к автоматизации деятельности по контролю качества трансформации данных на основе следующих процессов:

- записи и изменения доверенных значений параметров времени обработки данных в МЦП, утвержденных в договорах между ОС и предприятием-оператором МЦП и определяющих достигнутый уровень качества трансформации в ней данных;
- измерения фактических значений параметров времени трансформации данных;
- проведения сравнительной оценки фактических и доверенных значений параметров времени трансформации данных с определением и предоставлением информации об уровне качества трансформации лицам, принимающим решения.

В представленном исследовании решается задача формализации методического подхода к контролю качества трансформации данных в МЦП, удовлетворяющего изложенным выше требованиям.

2 Анализ процессов трансформации данных

Исходя из намерения формализации методического подхода к контролю качества трансформации данных произведен анализ входящего в МЦП и исходящего из МЦП процессов трансформации данных с учетом следующих факторов:

- намерений субъектов информационного взаимодействия посредством МЦП в отношении передачи добытых новых знаний одними субъектами и приобретения этих знаний другими субъектами [2];
- мирового опыта применения в цифровых транспортных сетях технологий трансформации данных, которые основаны на намерениях взаимодействующих субъектов, сетевом управлении при передаче данных и управлении информационно-технологическими услугами (см., например, [3–10]);

- методов оценки производительности при трансформации данных для ОС в облачных вычислительных структурах (см., например, [11–15]);
- функций и производительных ресурсов (функциональных, информационных, вычислительных, транспортных), используемых в МЦП в соответствии с регламентированными процедурами процессной технологии для трансформации данных, при которой для работы с каждой ОС закрепляют по одной прикладной программе с базой данных, а вычислительные ресурсы закрепляют за группами (подмножествами) ОС с одинаковыми доверенными значениями вероятностно-временных параметров.

Для наглядного отображения взаимосвязей ресурсов и функций разработана и представлена на рис. 1 схема производительных ресурсов МЦП.

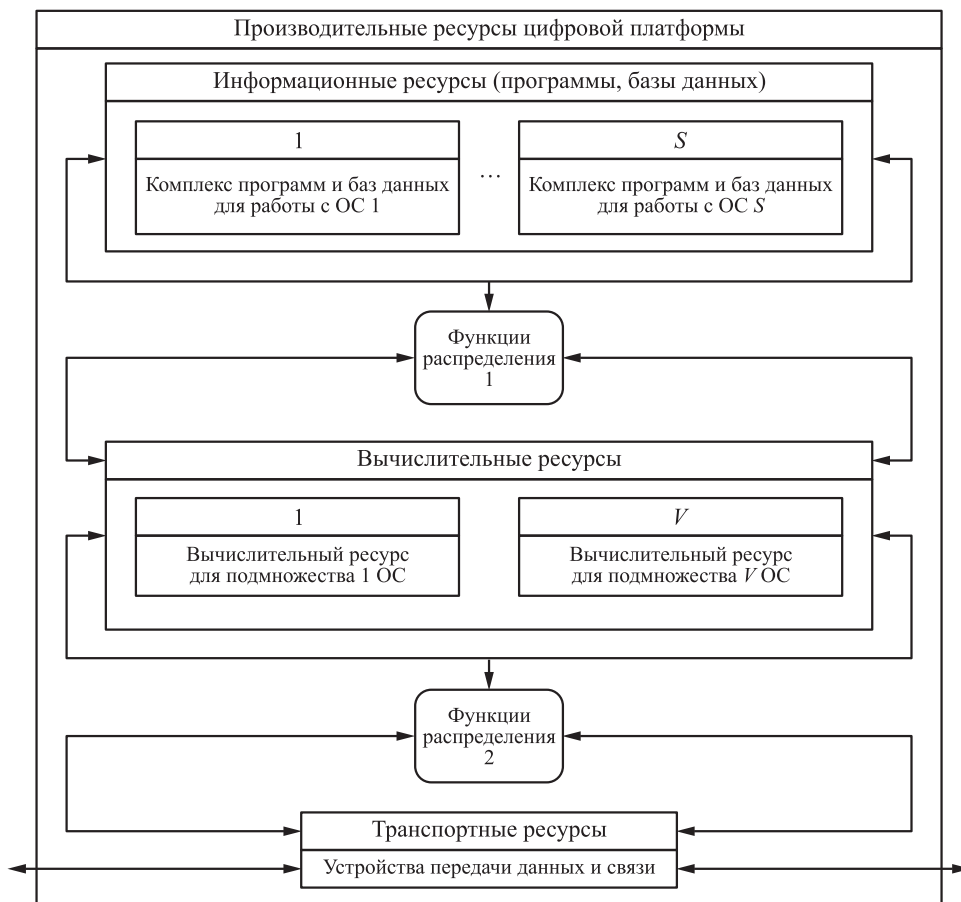


Рис. 1 Схема производительных ресурсов МЦП

На основании результатов проведенного анализа выявлено, что:

- время обработки трансформируемых данных носит динамический характер, который обусловлен зависимостью времени от информации, которую содержат данные;
- эта информация определяется намерениями субъектов взаимодействия и, как следствие, параметрами маршрутов для доставки информации в цифровой транспортной сети [1];
- важными для экономических агентов считаются все параметры своевременности доставки им этой информации, которая содержит новые знания, в том числе параметры трансформации данных на ресурсах МЦП.

3 Модели трансформации данных на ресурсах междисциплинарной цифровой платформы

С учетом приведенных выше соображений определен главный принцип контроля качества трансформации данных на групповых вычислительных ресурсах МЦП, который состоит в применении процедуры измерения в них длительности *интервалов занятости* с учетом динамических свойств времени обработки трансформируемых данных.

Разработана схема группового вычислительного ресурса МЦП (рис. 2), которая отображает изложенный выше подход.

При описании группового вычислительного ресурса МЦП (см. рис. 2) используются следующие параметры:

$P_N^k(j)$ — вероятность того, что требование — пакет данных, поступивший k -м по порядку в интервале занятости вычислительного ресурса длительностью в N пакетов данных — условных единиц, — застанет в нем j пакетов данных;

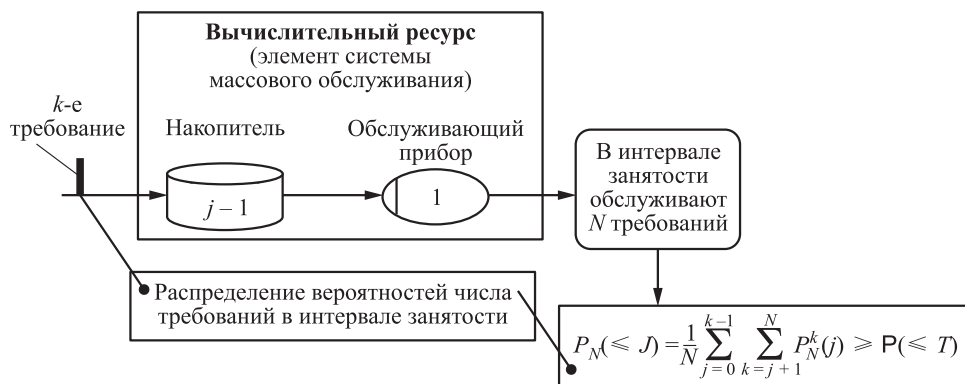


Рис. 2 Схема группового вычислительного ресурса МЦП

T — максимально допустимое (доверенное) значение времени ожидания пакетом данных обслуживания;

$P(\leq T)$ — минимально допустимое (доверенное) значение вероятности превышения времени T ;

$[J = (T/t_{cu})]$ — максимально допустимое значение времени ожидания пакетом данных обслуживания, выраженное числом условных единиц, где через $[x]$ обозначена целая часть числа x .

t_{cu} — длительность 1 условной единицы.

В МЦП образуются два типа трактов трансформации данных с использованием групповых вычислительных ресурсов.

Модель трансформации данных во входящем тракте представлена на рис. 3.

В модели трансформации данных во входящем тракте (см. рис. 3) выполняются следующие функции:

- функции распределения 2-го типа выполняют запись данных из транспортного ресурса в вычислительный ресурс в соответствии с адресом ОС, из которой данные переданы;
- вычислительный ресурс инициирует запуск функций распределения 1-го типа, используя код и базу данных той прикладной программы (функциональный и информационный ресурс), которая востребована для проведения трансформации над поступившими данными;
- функции распределения 1-го типа выполняют на вычислительном ресурсе необходимые операции, например по установке канала передачи данных в локальной сети между вычислительным комплексом, с одной стороны, и накопителем с программой и базой данных, с другой стороны;
- вычислительный ресурс обеспечивает выполнение операций программы, например по криптографическому дешифрованию данных, и передает данные в соответствующую базу данных (информационный ресурс).

Модель трансформации данных в исходящем тракте представляет собой зеркальное отображение модели трансформации данных во входящем тракте (см. рис. 3), при этом выполняются следующие функции:

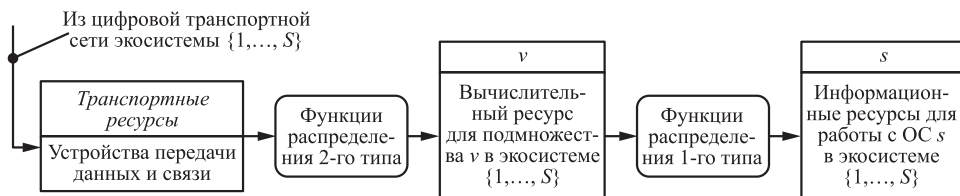


Рис. 3 Модель трансформации данных во входящем тракте

- соответствующая программа на информационных ресурсах инициирует запуск функций распределения 1-го типа, используя код того вычислительного ресурса, который востребован для проведения трансформации данных, например код ОС, в которую данные должны быть переданы;
- функции распределения 1-го типа выполняют процедуры, необходимые для выполнения операций на вычислительном ресурсе, например по установке канала передачи данных в локальной сети между вычислительным комплексом, с одной стороны, и накопителем с программой и базой данных, с другой стороны;
- вычислительный ресурс обеспечивает выполнение операций программы, например по криптографическому шифрованию данных, и инициирует запуск функций распределения 2-го типа, используя коды устройств передачи данных и связи, востребованные для передачи данных в соответствующую ОС;
- функции распределения 2-го типа выполняют запись данных из вычислительного ресурса последовательно в соответствующие устройства передачи данных и связи из состава транспортных ресурсов.

Транспортный ресурс осуществляет передачу данных в цифровую транспортную сеть и далее в соответствующую ОС.

4 Формальное описание методического подхода к контролю над трансформацией данных в междисциплинарной цифровой платформе

Методический подход к контролю над трансформацией данных в МЦП, исходя из представленной выше информации, формально можно представить как последовательность выполнения следующих шагов:

1. Заключают договоры с обязательствами оператора МЦП перед ОС — пользователями ее ресурсов — о соблюдении значений параметров времени передачи данных о новых знаниях экономическим агентам от их научных партнеров — времени, затрачиваемого на трансформацию этих данных в МЦП.
2. Определяют в качестве доверенных значений максимально допустимые значения времени T_v ожидания требованиями обслуживания на групповых вычислительных ресурсах, $v = 1, 2, \dots, V$ (см. рис. 2), и минимально допустимых значений вероятности $P_v (\leq T_v)$ превышения этого времени.
3. Преобразуют доверенные значения времени T_v для групповых вычислительных ресурсов в безразмерные значения $[J_v = T_v/t_{cu}^v]$, которые выражены числом условных единиц, причем:
 - формат данных соответствует разрядности вычислительного ресурса;

- время обработки одного формата (пакета) данных равно одной безразмерной условной единице (в процедурах измерений используются показатели времени — микро-, миллисекунды и др.).
- 4. Измеряют фактические значения длительности интервалов T_v^* занятости на групповых вычислительных ресурсах и преобразуют их к безразмерным значениям $[J_v^* = T_v^*/t_{cu}^v]$.
- 5. Проводят сравнительную оценку соответствующих доверенных и фактических безразмерных значений длительности интервалов занятости на групповых вычислительных ресурсах:

$$\left[J_v = \frac{T_v}{t_{cu}^v} \right] - \left[J_v^* = \frac{T_v^*}{t_{cu}^v} \right] = L_v .$$

Если $L_v \geq 0$, то классифицируют значение L_v как *доверенное значение* качества трансформации данных на групповом вычислительном ресурсе v и добавляют 1 (единицу) в базу данных доверенных значений МЦП.

Если $L_v < 0$, то переходят к шагу 6.

- 6. С использованием известной формулы [16, с. 13, формула (6)] проводят расчет вероятности $P_N^k(j)$ того, что пакет данных, поступивший k -м по порядку в интервале занятости вычислительного ресурса длительностью в N пакетов данных, застанет в нем j пакетов данных. Исходные данные для расчета:

$$j = J_v - 1; \quad k = N .$$

Если $P_N^k(j) \geq P_v(\leq T_v)$, то классифицируют значение L_v как *доверенное значение* качества трансформации данных на групповом вычислительном ресурсе v и добавляют 1 (единицу) в счетчик доверенного качества МЦП.

Если $P_N^k(j) < P_v(\leq T_v)$, то классифицируют значение L_v как *недостаточное значение* качества трансформации данных на групповом вычислительном ресурсе v добавляют 1 (единицу) в счетчик недостаточного качества МЦП.

- 7. Циклично с параметром ΔT (например, декада, месяц, квартал) по соглашению оператора МЦК и субъектов взаимодействия на основе показаний счетчиков доверенного (S_1) и недостаточного (S_2) качества МЦП проводят расчет уровня качества трансформации данных на вычислительных ресурсах:

если $[S_1/(S_1 + S_2)] \cdot 100\% \geq P_v(\leq T_v)$, то уровень качества — *доверенный*;

если $[S_1/(S_1 + S_2)] \cdot 100\% < P_v(\leq T_v)$, то уровень качества — *недостаточный*.

Формируют отчет и предоставляют лицам, принимающим решения.

Наличие расчетных процедур в предлагаемом методическом подходе к контролю качества позволяет получить более точные значения при оценке уровня качества трансформации данных в МЦП.

5 Заключение

Впервые поставлена, обоснована и решена научная задача о разработке методического подхода к контролю качества трансформации данных в МЦП, обеспечивающего проведение оценки степени соответствия фактических значений параметров времени трансформации данных их доверенным значениям, которые записаны в договорах субъектов с оператором цифровой платформы.

Разработана схема производительных ресурсов цифровой платформы, отображающая взаимосвязи функциональных, информационных, вычислительных и транспортных ресурсов, и типовая схема группового вычислительного ресурса, используемая при проведении оценки уровня качества трансформации данных.

Предложена двухуровневая процедура оценки качества трансформации данных, обеспечивающая повышение точности оценки качества, что стимулирует рост доверия субъектов и оператора цифровой платформы к получаемой ими информации о качестве обработки данных.

Предлагаемый методический подход к контролю качества трансформации данных может найти практическое применение в МЦП, предназначенных для организации и поддержки информационного взаимодействия между субъектами экономической, управленческой и научной деятельности при реализации задач федеральных проектов, выполняемых в соответствии с национальной программой «Цифровая экономика».

Литература

1. *Zatsarinnyy A. A., Shabanov A. P.* The software-controlled digital platform model for the transmission of new knowledge // Conference (International) on Modern Network Technologies Proceedings. — Piscataway, NJ, USA: IEEE, 2022. Art. 9960743. 8 p. doi: 10.1109/MoNeTec55448.2022.9960743.
2. *Xu J., Smyth H., Zerjav V.* Towards the dynamics of trust in the relationship between project-based firms and suppliers // *Int. J. Proj. Manag.*, 2021. Vol. 39. Iss. 1. P. 32–44. doi: 10.1016/j.ijproman.2020.09.005.
3. *Yazici A., Mishra A., Kontogiorgis P.* IT service management (ITSM) education and research: Global view // *Int. J. Eng. Educ.*, 2015. Vol. 31. Iss. 4. P. 1071–1080.
4. *Asikainen A.-L., Mangiarotti G.* Open innovation and growth in IT sector // *Serv. Bus.*, 2017. Vol. 11. P. 45–68. doi: 10.1007/s11628-015-0301-2.
5. *Lattemann C., Siemon D., Dorawa D., Redlich B.* Digitization of the design thinking process solving problems with geographically dispersed teams // *Design, user experience, and usability: Theory, methodology, and management* / Eds. A. Marcus, W. Wang. — Lecture notes in computer science ser. — Cham: Springer, 2017. Vol. 10288. P. 71–88. doi: 10.1007/978-3-319-58634-2_6.
6. *Shao G., Latif H., Martin C., Denno P.* Standards-based integration of advanced process control and optimization // *J. Industrial Information Integration*, 2019. Vol. 13. P. 1–12. doi: 10.1016/j.jii.2018.11.001.

7. *Semenov V., Ilyin D., Morozov S., Tarlapan O.* Effective consistency management for large-scale product data // *J. Industrial Information Integration*, 2019. Vol. 13. P. 13–21. doi: 10.1016/j.jii.2018.11.006.
8. *Потанов С. Е.* Исследование процесса передачи информации по виртуальным маршрутам в радиосети системы связи с подвижными объектами // *Теория и техника радиосвязи*, 2019. № 3. С. 11–23.
9. *Morshedzadeh I., Ng A. H. C., Jeusfeld M.* Managing manufacturing data and information in product lifecycle management systems considering changes and revisions // *Int. J. Product Lifecycle Management*, 2021. Vol. 13. Iss. 3. P. 244–264. doi: 10.1504/IJPLM.2021.118041.
10. *Auemhammer J., Bernard R.* The origin and evolution of Stanford University’s design thinking: From product design to design thinking in innovation management // *J. Prod. Innovat. Manag.*, 2021. Vol. 38. P. 623–644. doi: 10.1111/jpim.12594.
11. *Орехов С. Е., Артамонов Д. П., Иванов С. А.* Технологии облачных вычислений в интегрированных системах управления // *Инфокоммуникационные технологии*, 2020. Т. 18. № 4. С. 477–484. doi: 10.18469/ikt.2020.18.4.13.
12. *Volkov A. O.* Evaluation of cloud computing cluster performance // *T-Comm*, 2020. Vol. 14. Iss. 12. P. 72–79. doi: 10.36724/2072-8735-2020-14-12-72-79.
13. *Шейхгасанов Ш. К., Колотилов Ю. В.* Методы достижения энергоэффективности в облачных вычислениях // *Вестник Астраханского государственного технического университета. Сер. Управление, вычислительная техника и информатика*, 2020. № 2. С. 77–83. doi: 10.24143/2072-9502-2020-2-77-83.
14. *Файзуллин Р. В., Херинг Ш., Василенко К. А.* Модели оценки эффективности облачных технологий и туманных вычислений // *Моделирование, оптимизация и информационные технологии*, 2020. Т. 8. № 1. С. 30–31. doi: 10.26102/2310-6018/2020.28.1.025.
15. *Спицын А. А.* Алгоритм настройки параметров расписания облачных вычислений на основе оптимизации роя частиц // *Вестник Рязанского государственного радиотехнического университета*, 2021. Вып. 75. С. 44–52. doi: 10.21667/1995-4565-2021-75-44-52.
16. *Зацаринный А. А., Шабанов А. П.* Методологический подход к управлению качеством информации в сложных инфокоммуникационных проектах // *Системы и средства информатики*, 2011. Т. 21. № 2. С. 3–20.

Поступила в редакцию 07.05.23

METHODOLOGICAL APPROACH TO QUALITY CONTROL OF DATA TRANSFORMATION IN AN INTERDISCIPLINARY DIGITAL PLATFORM

A. A. Zatsarinny and A. P. Shabanov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The object of the research are the processes of data transformation in an interdisciplinary digital platform that provides information interaction between subjects of economic, managerial, and scientific activities. The results of the analysis of incoming and outgoing data transformation processes in the digital platform are presented. It is shown that the processing time of transformed data in computing resources depends on the information that these data contain and which is determined by the intentions of the subjects. It is revealed that for economic agents, the parameters of the timeliness of providing information about new knowledge which is transmitted to them by scientific partners are important. The parameters of data processing time in computing resources are classified as data transformation quality parameters which are approved in the contracts of subjects with the platform operator and are taken into account in the requirements for quality control processes. For the first time, a methodological approach to quality control is proposed based on a comparative assessment of the degree of compliance of the actual values of the time parameters of data transformation with their trusted values which are approved in contracts. The practical significance of the methodological approach is to provide the functions of an objective tool for monitoring compliance with the contractual obligations of the operator of the digital platform to the interacting entities.

Keywords: digital platform; data transformation; information interaction; timeliness; trusted control

DOI: 10.14357/08696527230311

EDN: MEONXL

References

1. Zatsarinnyy, A. A., and A. P. Shabanov. 2022. The software-controlled digital platform model for the transmission of new knowledge. *Conference (International) on Modern Network Technologies Proceedings*. Piscataway, NJ: IEEE. 9960743. 8 p. doi: 10.1109/MoNeTec55448.2022.9960743.
2. Xu, J., H. Smyth, and V. Zerjav. 2021. Towards the dynamics of trust in the relationship between project-based firms and suppliers. *Int. J. Proj. Manag.* 39(1):32–44. doi: 10.1016/j.ijproman.2020.09.005.
3. Yazici, A., A. Mishra, and P. Kontogiorgis. 2015. IT service management (ITSM) education and research: Global view. *Int. J. Eng. Educ.* 31(4):1071–1080.

4. Asikainen, A.-L., and G. Mangiarotti. 2017. Open innovation and growth in IT sector. *Serv. Bus.* 11:45–68. doi: 10.1007/s11628-015-0301-2.
5. Lattemann, C., D. Siemon, D. Dorawa, and B. Redlich. 2017. Digitization of the design thinking process solving problems with geographically dispersed teams. *Design, user experience, and usability: Theory, methodology, and management*. Eds. A. Marcus and W. Wang. Lecture notes in computer science ser. Cham: Springer. 10288:71–88. doi: 10.1007/978-3-319-58634-2_6.
6. Shao, G., H. Latif, C. Martin, and P. Denno. 2019. Standards-based integration of advanced process control and optimization. *J. Industrial Information Integration* 13:1–12. doi: 10.1016/j.jii.2018.11.001.
7. Semenov, V., D. Ilyin, S. Morozov, and O. Tarlapan. 2019. Effective consistency management for large-scale product data. *J. Industrial Information Integration* 13:13–21. doi: 10.1016/j.jii.2018.11.006.
8. Potapov, S. E. 2019. Issledovanie protsessa peredachi informatsii po virtual'nym marshrutam v radioseti sistemy svyazi s podvizhnymi ob'ektami [Research of process of information transfer on virtual routes in the radio network of the communication system with mobile objects]. *Teoriya i tekhnika radiosvyazi* [Radio Communication Theory and Technology] 3:11–23.
9. Morshedzadeh, I., A. H. C. Ng, and M. Jeusfeld. 2021. Managing manufacturing data and information in product lifecycle management systems considering changes and revisions. *Int. J. Product Lifecycle Management* 13(3):244–264. doi: 10.1504/IJPLM.2021.118041.
10. Auemhammer, J., and R. Bernard. 2021. The origin and evolution of Stanford University's design thinking: From product design to design thinking in innovation management. *J. Prod. Innovat. Manag.* 38(6):623–644. doi: 10.1111/jpim.12594.
11. Orekhov, S. E., D. P. Artamonov, and S. A. Ivanov. 2020. Tekhnologii oblachnykh vychisleniy v integrirovannykh sistemakh upravleniya [Cloud computing technologies in integrated management systems]. *Infokommunikatsionnye tekhnologii* [Infocommunication Technologies] 18(4):477–484. doi: 10.18469/ikt.2020.18.4.13.
12. Volkov, A. O. 2020. Evaluation of cloud computing cluster performance. *T-Comm* 14(12):72–79. doi: 10.36724/2072-8735-2020-14-12-72-79.
13. Sheykhgasanov, Sh. K., and Yu. V. Kolotilov. 2020. Metody dostizheniya energoeffektivnosti v oblachnykh vychisleniyakh [Methods for achieving energy efficiency in cloud computing]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Ser. Upravlenie, vychislitel'naya tekhnika i informatika* [Bulletin of Astrakhan State Technical University. Ser. Management, Computer Science, and Informatics] 2:77–83. doi: 10.24143/2072-9502-2020-2-77-83.
14. Fayzullin, R. V., Sh. Khering, and K. A. Vasilenko. 2020. Modeli otsenki effektivnosti oblachnykh tekhnologiy i tumannykh vychisleniy [Models of evaluations of the cloud technology and fog computing]. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [Modeling, Optimization, and Information Technology] 8(1):30–31. doi: 10.26102/2310-6018/2020.28.1.025.
15. Spitsyn, A. A. 2021. Algoritm nastroyki parametrov raspisaniya oblachnykh vychisleniy na osnove optimizatsii roya chastits [Algorithm for configuring cloud computing schedule parameters based on particle swarm optimization]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta* [Bulletin of the Ryazan State Radio Engineering University] 75:44–52. doi: 10.21667/1995-4565-2021-75-44-52.

16. Zatsarinny, A. A., and A. P. Shabanov. 2011. Metodologicheskiy podkhod k upravleniyu kachestvom informatsii v slozhnykh infokommunikatsionnykh proektakh [Methodological approach to quality management in complex communication projects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21(2):3–20.

Received May 7, 2023

Contributors

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; AZatsarinny@ipiran.ru

Shabanov Alexander P. (b. 1949) — Doctor of Science in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; apshabanov@mail.ru

ТЕОРИЯ S-СИМВОЛОВ: СЕТЕВОЙ ТАБС-РЕШАТЕЛЬ S-ЗАДАЧ

В. Д. Ильин¹

Аннотация: Теория S-символов представляет собой расширенное обобщение теории S-моделирования; рассматривается как часть методологического обеспечения разработок систем искусственного интеллекта в S-среде (включая системы знаний, системы S-моделирования задач и конструирования программ и др.). S-среда, основанная на взаимосвязанных системах S-(символов, кодов, сигналов), служит инфраструктурным основанием реализации информационных технологий различного назначения. В статье представлена третья часть (из четырех) описания теории. Приведены определения понятий многослойной таблицы (табса) и табс-структур, используемых для представления S-объектов. Дано описание основ табс-конструирования S-объектов. Основная часть статьи посвящена сетевому табс-решателю произвольных задач на основе системы знаний об S-задачах. Методология его построения рассматривается как вклад в методологическое обеспечение разработки систем цифровизации различных видов деятельности.

Ключевые слова: теория S-символов; S-задача; табс; сетевой табс-решатель S-задач; автоматизация разработки систем цифровизации

DOI: 10.14357/08696527230312

EDN: EJHIQJ

1 Введение

Уже более полувека не прекращаются попытки найти эффективный ответ на вопрос, как автоматизировать процессы программирования и решения задач из произвольных предметных областей.

На ранних этапах (в 1970–1980-х гг.) особенно популярными были идеи автоматического синтеза программ [1]. Сторонники этих идей рассматривали программы и спецификации задач как математические сущности, полагая, что первые могут быть получены путем формальных преобразований вторых. Одной из известных попыток на этом направлении стала система ПРИЗ [2, 3].

Но тем, кто имеет успешный опыт разработки нетривиальных программных продуктов, хорошо известно, что спецификация нетривиальной задачи и программа соотносятся так же, как проектная документация (в инженерном деле) и изделие, которое предстоит изготовить.

Запись и выделение фрагментов текста. Для записи формул и выделения определений, замечаний и примеров в составе *S-сообщений* используются средства языка *TSM-комплекса* (*TSM: textual symbolic modeling*)².

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vdilyin@yandex.ru

²Символьное моделирование (S-моделирование) // Энциклопедия Руниверсалис. [https://руни.рф/index.php/Символьное_моделирование_\(S-моделирование\)](https://руни.рф/index.php/Символьное_моделирование_(S-моделирование)).

Язык *TSM* имеет средства записи формализованных выражений (без применения редакторов формул), выделения частей *S*-сообщений и замены выбранными сокращениями часто повторяющихся фрагментов. Для выполнения описаний достаточно стандартной клавиатуры и набора средств, имеющихся в составе текстовых редакторов пакетов *LibreOffice*, *OpenOffice* или др.

В статье применены следующие *TSM*-средства выделения фрагментов текста:

□ ⟨фрагмент описания⟩ □ \approx определение (здесь и далее символ \approx заменяет слово «означает»);

◇ ⟨фрагмент описания⟩ ◇ \approx замечание;

○ ⟨фрагмент описания⟩ ○ \approx пример.

Курсивом (и полужирным) выделены первые вхождения названий понятий и фрагменты описания, к которым автор хочет привлечь внимание.

2 Табс и табс-структура

Идея изобретения *табса* и *табс-структуры* сформировалась в процессе поиска системы форм, удобной для интерактивного описания *множеств понятий и заданных на них семейств S-задачных отношений* [4]. При этом искомая позиционная система должна была быть удобной и эффективной для представления на ее основе всех структур данных в процессах *S*-задачного конструирования.

□ *Табс* — это трехмерная (многослойная) таблица, имеющая два типа клеток для связи с другими такими таблицами. Первый тип — *указатель на табс* (*tabs-pointer*), второй — *вложенный табс* (*tabs-nested*). Значениями типа *tabs-pointer* служат табс-адреса. *Табс-адрес* (*tabs-adress*) задается именем табса и тремя координатами: первая — *номер табс-строку* (*tabs-row*), вторая — *табс-столбца* (*tabs-column*) и третья — *табс-слоя* (*tabs-layer*). Значениями типа *tabs-nested* служат табсы. Попадание в клетку типа *tabs-nested* делает доступным вложенный в эту клетку табс. □

□ *Табс-структура* — множество табсов и семейство заданных на нем отношений. Отношения реализуются посредством типов *tabs-nested* и *tabs-pointer*. □

◇ Существование типа *tabs-nested* освобождает от необходимости во всех случаях именовать табсы. Возможность создавать вложенные табсы полезна и для агрегирования, и для декомпозиции. ◇ ○ Клетку типа *tabs-nested* можно рассматривать как некоторый агрегат, представляющий фрагмент памяти конструируемого *табс-решателя S-задач* (*tabs-solver*), а сам вложенный табс при этом рассматривать как декомпозицию упомянутого фрагмента. ○

3 Сетевой табс-решатель S-задач (*tabs-solver*)

□ *Tabs-solver* представляет собой сеть решателей *S*-задач, в которой каждый узел *tabs-solver^N* может взаимодействовать с любым другим узлом этой сети. □

Каждый $tabs-solver^N$ имеет цифрового двойника [5] и специализируется на определенном классе задач из заданной предметной области (\bigcirc это могут быть задачи построения системы знаний об S-задачах, автоматизированного конструирования программ или др. \bigcirc). Цифровые двойники служат средством совершенствования функционирования отдельных $tabs-solver^N$ и всего $tabs-solver$.

\square **S-задача** $\approx \{Formul, Rulsys, Alg, Prog\}$, где *Formul* — постановка S-задачи; *Rulsys* — множество систем *обязательных* и *ориентирующих требований* к решению S-задачи, поставленных в соответствие *Formul*; *Alg* — объединение множеств алгоритмов, каждое из которых соответствует одному элементу из *Rulsys*; *Prog* — объединение множеств *S-программ*, каждое из которых поставлено в соответствие одному из элементов *Alg*. Постановка S-задачи $Formul \approx \{Mem, Rel\}$, где *Mem* — множество понятий задачи, на котором задано разбиение $Mem = Inp \cup Out$ ($Inp \cap Out = 0$) и совокупность *Rel* связей между понятиями, определяющая бинарное отношение $Rel < Inp * Out$. Множество *Mem* называем *памятью задачи*, а *Inp* и *Out* — ее *входом* и *выходом*, значения которых предполагается соответственно задавать и искать. \square

В общем случае множества *Rulsys*, *Alg* и *Prog* могут быть пустыми: число их элементов зависит от степени изученности S-задачи. Для каждого элемента из *Rulsys*, *Alg* и *Prog* задано описание применения. Описания применения элементов *Rulsys* включают спецификацию типа *решателя задачи*, требование к информационной безопасности и др. Описания применения элементов из *Alg* включают данные о допустимых режимах работы решателя S-задачи, о требованиях к полученному результату и др. Описания применения S-программ включают данные о языках программирования, операционных системах и др. [6, 7].

Будем рассматривать $tabs-solver^N$ с двух точек зрения.

Во-первых, как информационный автомат, взаимодействующий с себе подобными автоматами и с *элементами окружения* (далее — *enels*) путем посылки S-сообщений. Такое представление нужно, чтобы, не вникая в механизмы решения задач, работать с $tabs-solver^N$ как с активным S-объектом, представляющим собой вершину табс-представленного S-графа [7].

Второй облик $tabs-solver^N$ используем, чтобы раскрыть его предназначение как средства решения S-задач определенной предметной области. Реализация S-модели предметной области, интерпретация запросов на S-модели — такие вопросы будем рассматривать на основе второго представления.

\square **Информационная модель $tabs-solver^N$** — это S-машина с конечным множеством состояний *S*, обладающая *целью* (понимаемой как множество S-задач, разрешимых на задачном S-графе *G* [7]). \square

Множество *T* пар состояний, принадлежащих декартову произведению $S * S$, задает отношение допустимых переходов. Пара (s_i, s_j) принадлежит *T*, если из состояния s_i разрешен переход в состояние s_j ; $i, j = 1, \dots, n_S$ (n_S — число состояний). S-граф *G* служит формализованным представлением *задачной области*, рассчитанным на поиск *разрешающих структур* и формализацию знаний

об S-задачах [7]. Множество вершин S-графа составлено из S-объектов. Каждая вершина S-графа имеет табс-память (представленную памятью S-задачи или S-задачной области). Ребро S-графа — пара вершин с непустым пересечением по табс-памяти. Нагрузка ребра определяется множеством всех пар элементов табс-памяти, входящих в это пересечение [7].

□ Называем $tabs-solver^N$ вычислительным, если каждому s_i из множества S состояний алгоритмически поставлено в соответствие не более одного s_j ($i \neq j$). Если же каждому s_i из множества S может быть поставлено в соответствие более одного s_j ($i \neq j$), $tabs-solver^N$ называем исчислительным. □

Поведение $tabs-solver^N$ определяется следующими требованиями:

- принимать и отправлять S-сообщения, взаимодействуя с другими $tabs-solver^N$ и $enels$;
- интерпретировать S-сообщения, выдавая сведения об ошибках и допуская коррекцию принятых S-сообщений;
- выполнять работу, запрошенную в S-сообщении, допуская диалог, паузы и отмену;
- оформлять результат выполненной работы для отправки корреспонденту;
- образовывать объединения $tabs-solver^N$, предназначенные для решения комплексов S-задач.

$tabs-solver^N$ работает как S-машина для обработки S-сообщений безотносительно к содержанию работы, выполняемой в соответствии с поступившим запросом. Такой взгляд на $tabs-solver^N$ соответствует намерению исследовать проблему создания сети табс-решателей S-задач.

$tabs-solver^N$ как S-машина для обработки S-сообщений состоит из следующих компонент, составляющих конвейер (которые могут работать асинхронно):

- $ts-before$ — инициализация, запуск, обучение;
- $ts-start$ — прием, буферизация, интерпретация S-сообщений;
- $ts-step$ — шаг в процессе выполнения работы, запрошенной в S-сообщении;
- $ts-end$ — завершение работы с текущим S-сообщением;
- $ts-after$ — завершение работы.

В то время когда компонента $ts-before$ занята (○ обучением пользователя ○), компонента $ts-start$ может интерпретировать очередное S-сообщение, а $ts-step$ — выполнять работу, запрошенную в предыдущем S-сообщении.

Каждая компонента $tabs-solver^N$ может находиться в одном из состояний, принадлежащих множеству

$$State = \{head, tail, dial, intr, paus, canc, resl, fail, null\},$$

где $head$ — состояние готовности; $tail$ — работа закончена; $dial$ — взаимодействие с $enels$; $intr$ — взаимодействие с другими $tabs-solver^N$; $paus$ — пауза;

nc — отмена; *resl* — результат получен; *fail* — неудача: результат получить не удалось; *null* — состояние, требующее непредусмотренного вмешательства отправителя сообщения (*enel* или *tabs-solver^N*).

Отношение перехода из одного состояния в другое задано для каждой компоненты соответствующей таблицей перехода состояний. Таблицы для всех компонент имеют одинаковую структуру: число строк и число столбцов равны числу состояний. Строки и столбцы имеют те же имена, что и соответствующие состояния. Если в клетке на пересечении строки st_i и столбца st_j стоит буква f , это означает, что переход из состояния st_i в состояние st_j невозможен. Если в клетке размещена ссылка на правило *rule*, определяющее переход ($st_i \rightarrow st_j$), то это означает, что разрешенный переход выполняется в соответствии с этим правилом.

Компоненты взаимодействуют в соответствии с правилами, хранящимися в табс-базе правил взаимодействия компонент *tabs-solver^N*.

□ ***Tabs-solver^N* как решатель S-задач** из некоторой предметной области располагает средствами для интерпретации S-сообщений и выполнения запрашиваемых работ.

В состав *tabs-solver^N*, решающего S-задачи из заданной предметной области, входят:

- анализатор сообщений (запросов на решение S-задач);
- построитель разрешающих структур на S-графе;
- *система управления табс-базами данных* (СУТБД);
- табс-ориентированный процессор ввода–вывода;
- средства расширения множества разрешимых S-задач;
- функциональная компонента. □

Язык сообщений лаконичен (необязательные синтаксические конструкции заключены в фигурные скобки):

$$\begin{aligned} \langle \text{сообщение} \rangle &::= \langle \text{конверт} \rangle \{ \langle \text{содержимое} \rangle \} \\ \langle \text{конверт} \rangle &::= \langle \text{получатель} \rangle, \langle \text{отправитель} \rangle \\ \langle \text{содержимое} \rangle &::= \text{найти} \langle \text{выход} \rangle \text{по} \langle \text{вход} \rangle \{ \text{при} \langle \text{зависимость выход-вход} \rangle \} \\ &\quad \langle \text{разрешающая структура} \rangle \langle \text{адрес входа} \rangle \langle \text{адрес выхода} \rangle \} \end{aligned}$$

Получатель и отправитель представлены именами *tabs-solver^N* с указанием табс-координат портов S-сообщений. Выход и вход задаются именами понятий предметной области. Зависимость выход–вход задается в тех случаях, когда запрос представлен в виде формулировки S-задачи.

Разрешающая структура представлена именем; адрес входа указывает место, откуда следует прочитать исходные данные; адрес выхода — куда следует записать полученный результат. Если один из адресов (или оба) не указан, это означает, что работа выполняется с использованием входных и выходных портов (для данных и результатов).

Обнаружив сообщение во входном порте сообщений, анализатор начинает синтаксический разбор. Если сообщение синтаксически верно, оно направляется строителю разрешающих структур. В противном случае отправителю выдаются сведения об ошибке (с предложением исправить сообщение). Когда получен искомым результат, он направляется отправителю по адресу, указанному на конверте.

Построитель разрешающих структур отыскивает на S-графе разрешающую структуру, позволяющую по заданному входу найти искомым выход.

Система управления табс-базами данных играет роль памяти *tabs-solver^N*. Все функции, свойственные СУТБД, присущи табс-памяти *tabs-solver^N* [4].

Формоориентированный процессор ввода–вывода представляет собой объединение структурно-ориентированных редакторов ввода, формирователей отчетов и средств вывода во внешнюю (по отношению к *tabs-solver^N*) среду.

Расширение множества разрешимых задач достигается увеличением числа вершин и ребер S-графа. Такая работа выполняется в специальном режиме «расширение» на копии расширяемого S-графа.

Целевое назначение *tabs-solver^N* представляет его функциональная компонента. Содержание задач, служащих вершинами S-графа, связи по табс-памяти между S-задачами, представленные ребрами графа, — все это определяет S-задачный облик функциональной компоненты.

4 Заключение

Представлена третья из четырех частей *теории S-символов*, рассматриваемой как часть методологического обеспечения разработок систем искусственного интеллекта (○ систем знаний, систем автоматизации программирования и др. ○) в цифровой среде. Предложенная S-модель сетевого табс-решателя на основе системы знаний об S-задачах рассматривается как вклад в методологическое обеспечение разработки систем цифровизации различных видов деятельности.

Литература

1. Manna Z., Waldinger R. Knowledge and reasoning in program synthesis // *Artif. Intell.*, 1975. Vol. 6. No. 2. P. 175–208. doi: 10.1016/0004-3702(75)90008-9.
2. Тыгу Э. Х. Концептуальное программирование. — М.: Наука, 1984. 256 с.
3. Кахро М. И., Калья А. П., Тыгу Э. Х. Инструментальная система программирования ЕС ЭВМ (ПРИЗ). — М.: Финансы и статистика, 1988. 181 с.
4. Ильин В. Д. Система порождения программ. — М.: Наука, 1989. 264 с.
5. The digital twin. — General Electric, 2018. https://www.ge.com/digital/sites/default/files/The-Digital-Twin_Compressing-Time-to-Value-for-Digital-Industrial-Companies.pdf.
6. Ильин В. Д. Теория S-символов: концептуальные основания // *Системы и средства информатики*, 2023. Т. 33. № 1. С. 126–134. doi: 10.14357/08696527230112.

7. Ильин В. Д. Теория S-символов: формализация знаний об S-задачах // Системы и средства информатики, 2023. Т. 33. № 2. С. 124–131. doi: 10.14357/08696527230212.

Поступила в редакцию 23.06.23

THEORY OF S-SYMBOLS: NETWORK TABS-SOLVER OF S-PROBLEMS

V. D. Ilyin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The theory of S-symbols is an extended generalization of the theory of S-modeling. It is considered as a part of the methodological support for the development of artificial intelligence systems in the S-environment (including knowledge systems, systems of S-modeling of problems and program design, etc.). The S-environment based on interconnected systems of S-(symbols, codes, signals) serves as the infrastructural basis for the implementation of information technologies for various purposes. The article presents the third part (out of four) of the description of the theory. Definitions of the concepts of a multilayer table (tabs) and tabs structures used to represent S-objects are provided. A description of the basics of tabs-construction of S-objects is given. The main part of the article is devoted to a network tabs solver of arbitrary problems based on a system of knowledge about S-problems. The methodology of its construction is considered as a contribution to the methodological support of the development of digitalization systems for various types of activities.

Keywords: theory of S-symbols; S-problem; tabs; network tabs-solver of S-problems; automation of digitalization systems development

DOI: 10.14357/08696527230312

EDN: EJHIQJ

References

1. Manna, Z., and R. Waldinger. 1975. Knowledge and reasoning in program synthesis. *Artif. Intell.* 6(2):175–208. doi: 10.1016/0004-3702(75)90008-9.
2. Tyugu, E. Kh. 1984. *Kontseptual'noe programmirovaniye* [Conceptual programming]. Moscow: Nauka. 256 p.
3. Kakhro, M. I., A. P. Kal'ya, and E. Kh. Tyugu. 1988. *Instrumental'naya sistema programmirovaniya ES EVM (PRIZ)* [The instrumental programming system for ES computer (PRIZ)]. Moscow: Finance and Statistics. 181 p.
4. Ilyin, V. D. 1989. *Sistema porozhdeniya programm* [The system of program generating]. Moscow: Nauka. 264 p.
5. General Electric. 2018. The digital twin. Available at: https://www.ge.com/digital/sites/default/files/The-Digital-Twin_Compressing-Time-to-Value-for-Digital-Industrial-Companies.pdf (accessed August 15, 2023).

6. Ilyin, V. D. 2023. Teoriya S-simvolov: kontseptual'nye osnovaniya [Theory of S-symbols: Conceptual foundations]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(1):126–134. doi: 10.14357/08696527230112.
7. Ilyin, V. D. 2023. Teoriya S-simvolov: formalizatsiya znaniy ob S-zadachakh [Theory of S-symbols: Formalization of knowledge about S-problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):124–131. doi: 10.14357/08696527230212.

Received June 23, 2023

Contributor

Ilyin Vladimir D. (b. 1937) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vdilyin@yandex.ru

ОЧИСТКА ДАННЫХ В ТЕХНОЛОГИИ ПОДДЕРЖКИ КОНКРЕТНО-ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЙ

И. М. Адамович¹, О. И. Волков²

Аннотация: Статья продолжает серию работ, посвященных технологии поддержки конкретно-исторических исследований (ПКИИ). Технология построена на принципах сотворчества и краудсорсинга и ориентирована на широкий круг не относящихся к профессиональным историкам и биографам пользователей. Показана целесообразность расширения перечня задач конкретно-исторического исследования, решаемых в рамках описанной технологии с применением методов машинного обучения. Отмечена особая важность подготовки данных в связи с фрагментарностью и противоречивостью конкретно-исторической информации. Данная статья посвящена специфике очистки конкретно-исторических данных и анализу возможности применения с этой целью механизмов и алгоритмов, уже интегрированных в технологию. Перечислены основные направления, по которым проводится очистка данных. Для каждого направления выявлены подходящие, уже включенные в технологию инструменты. Особое внимание уделено инструментам устранения несогласованности. Перечислены этапы очистки данных и приведена схема взаимодействия всех описанных в статье механизмов и алгоритмов.

Ключевые слова: конкретно-историческое исследование; распределенная технология; машинное обучение; очистка данных; несогласованность данных

DOI: 10.14357/08696527230313

EDN: DDSTXY

1 Введение

Поддержка конкретно-исторических исследований стала одной из актуальных задач современности, что обусловлено вовлечением в исследовательский процесс не только членов профессионального исторического сообщества, но и самых широких слоев непрофессионалов в связи со все возрастающим интересом к частной, семейной истории [1].

В [2, 3] описана разработанная в ФИЦ ИУ РАН распределенная технология ПКИИ, основанная на принципах краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий). Данные в этой технологии организованы в форме семантической сети. Узлы сети представляют собой именованные универсальные классы объектов. Факты задаются

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Adam@amsd.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Volkov@amsd.com

значениями экземпляров классов и связями между ними. Связи наследуются из сети классов [4]. Специфика данных технологии ПККИИ состоит в том, что исследователь при проведении конкретно-исторического исследования, как правило, имеет дело не с объективным знанием об объекте исследования, а с его отражением в исторических документах, которые обладают свойством фрагментарности и противоречивости и не существуют в отрыве от своих метаданных, т. е. информации о документе-источнике.

В процессе развития технологии в нее был включен ряд механизмов поддержки автоматизации таких подзадач конкретно-исторических исследований, как:

- определение потенциально перспективных направлений исследования за счет автоматического формирования гипотез о значении информационных лагун в наборе конкретно-исторических данных с опорой на процедуру импутации данных (восстановления пропущенных данных на основании анализа остальных данных в наборе), для которой разработан ряд методов [5];
- поиск аномалий в конкретно-исторических данных посредством кластеризации.

Данные механизмы были реализованы на базе алгоритма машинного обучения CHAID (chi-square automatic interaction detection). Допустимость его применения обуславливается возможностью сформировать выборку однотипных данных. В условиях технологии ПККИИ эта возможность не выглядит очевидной, поскольку факты, собираемые вручную одним пользователем технологии в рамках одного конкретно-исторического исследования, как правило, по своей природе единичны и разрознены и не могут составить достаточную по объему однородную выборку. Но, поскольку технология ПККИИ является распределенной и содержит средства автоматической обработки исторических источников, база данных (БД) технологии вполне может содержать множественные однородные факты в количестве, вполне достаточном для их обработки алгоритмами машинного обучения. Источниками множественных фактов в технологии ПККИИ могут служить:

- (1) результаты автоматической обработки нарративов.

Технология ПККИИ имеет в своем составе средства автоматического извлечения фактов из текстов историко-биографической направленности на естественном языке повествовательного характера (нарративных источников);

- (2) исторические реестры номинативного типа.

Номинативные (содержащие персональную информацию с указанием имени собственного) исторические источники можно разделить на одиночные, содержащие информацию об одной персоне или одной семье (метрические выписи, посемейные списки и т. д.), и реестры, содержащие информацию о множестве персон или семей (метрические книги и т. д.). Обработка

исследователем номинативных исторических реестров как вручную, так и с использованием интегрированных в технологию средств автоматизации может служить источником множественных однотипных фактов;

(3) разрозненные одиночные номинативные источники.

Поскольку технология ПКИИ была построена на принципах сотворчества и краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий), ее данные, организованные в форме семантической сети, пополняются независимо различными исследователями. Информация, найденная одним пользователем в рамках своего исследования, может быть доступна другим пользователям. Из этого следует, что разрозненные одиночные номинативные источники, а также фрагменты реестров номинативного типа, обработанные разными пользователями, могут стать источниками множественных однотипных фактов.

Но на вышеперечисленных подзадачах конкретно-исторического исследования возможности применения алгоритмов машинного обучения в технологии ПКИИ не исчерпываются. Так, как показано в [6], в конкретно-исторических исследованиях применяются следующие методы и приемы работы с данными:

- извлечение и группировка изучаемых объектов; сравнение групп объектов; получение сводных данных об изучаемых группах объектов;
- классификация изучаемых объектов;
- формирование хронологических рядов; периодизация (деление процессов развития на отличающиеся друг от друга периоды).

Задачи группировки и периодизации могут быть решены средствами кластеризации, что наряду с классификацией относится к типичным задачам машинного обучения [7]. Из этого следует целесообразность расширения перечня задач конкретно-исторического исследования, решаемых в рамках технологии ПКИИ с применением методов машинного обучения, и поиска для них оптимальных алгоритмов. Но подготовка данных для их дальнейшей обработки формальными алгоритмами анализа (классификации, кластеризации, регрессии и др.) имеет важное значение, поскольку «грязные» данные приводят к неправильным результатам и вводящей в заблуждение статистике [8]. Под грязными данными понимаются дублирующиеся, отсутствующие, несогласованные и просто ошибочные данные, включая те, которые не нарушают никаких ограничений, но все равно являются неправильными или непригодными для использования. Вот почему процесс обнаружения и исправления грязных данных, называемый очисткой данных (data cleansing, data cleaning, data scrubbing или data wrangling), становится обязательным условием любой задачи обработки данных [9].

Данная статья посвящена специфике очистки конкретно-исторических данных и анализу возможности применения с этой целью механизмов и алгоритмов, уже интегрированных в технологию ПКИИ.

2 Очистка данных

К общим операциям по очистке данных можно отнести [10]:

- заполнение пропущенных значений;
- сглаживание зашумленных данных;
- устранение несогласованности;
- удаление выбросов.

Перечисленные задачи особо актуальны в технологии ПКИИ, поскольку историко-биографические факты обладают фрагментарностью и противоречивостью [11], а основные причины этого кроются в следующих негативных особенностях номинативных исторических источников [12]:

- обилие несоответствий и ошибок в метрических записях;
- отсутствие в метрических книгах прихода фамилий крестьян вплоть до середины 1880-х гг.;
- большая вариативность в написании имен;
- указание в записях возраста с точностью до года;
- нередко встречающиеся искажения возраста, в том числе сознательные;
- частое неупоминание сведений о женщинах;
- частое неупоминание сведений о детях;
- частое смешение детей, рожденных в разных браках, и т. д.

В технологии ПКИИ уже предусмотрены некоторые средства, позволяющие выполнить очистку данных. Так, для заполнения пропущенных значений предусмотрена процедура импутации данных, основанная на алгоритме SNAID, включенная в технологию в рамках механизма формирования гипотез для определения перспективных направлений поиска [13]. Для борьбы с шумами в данных и выбросами могут быть использованы алгоритмы кластеризации [14], включенные в технологию в рамках механизма выявления аномалий в конкретно-исторических данных [15].

Наиболее остро в технологии ПКИИ стоит проблема несогласованности данных, вытекающая из приведенной выше специфики номинативных исторических источников. Под несогласованностью данных обычно понимается такое негативное следствие избыточности данных, как наличие конфликтующих между собой записей или расхождений. Конфликт возможен в формате данных и в значениях данных.

Формат данных в технологии ПКИИ задается форматами в сети классов, и соответствие форматов экземпляров обеспечивается их наследованием от одного класса. Но разночтения в значениях данных становятся серьезной проблемой. Такие разночтения возникают вследствие получения информации об одной персоне из различных источников, каждый из которых по разным причинам может

содержать недостоверные сведения, сведения, отражающие реальные изменения в наименованиях различных объектов или вариативность их написания, а также какие-то уточнения. Появление в документах недостоверных сведений может быть вызвано как непреднамеренными ошибками при составлении документов, так и преднамеренными искажениями. Так, в некоторых документах крестьян конца XIX – начала XX вв. мог быть указан преднамеренно заниженный возраст с целью избежать воинской повинности во время военных действий. Реальные изменения наименования могли быть вызваны изменениями границ губерний или переименованиями населенных пунктов. Вариативность написания особо часто проявляется в именах людей. В качестве примера можно привести наименование одной и той же женщины в разных документах Лукерьей и Гликерией, что хоть и затрудняет ее автоматическую идентификацию, но не исключает полностью, поскольку это две принятые формы одного имени. Пример женщины, названной в одном документе Марфой, а в другом — Маврой, более сложен, поскольку Марфа и Мавра — два разных имени и путаница вызвана исключительно сходством их звучания. Выявление таких несоответствий — сложная задача, но в технологии ПКИИ предусмотрены механизмы, способные существенно облегчить ее решение.

3 Механизмы устранения несогласованности данных

В технологии ПКИИ предусмотрен ряд механизмов, позволяющих осуществить устранение несогласованности данных.

3.1 Нормали

Существуют объективные зависимости, увязывающие объекты исторической реальности с их динамическими характеристиками (атрибутами). Эти зависимости регламентируются естественными закономерностями, определяющими допустимые последовательности событий и задающими некий шаблон, сценарий или набор ограничений, законами природы, законами и нормативными документами, традициями. Формулировки этих регламентирующих закономерностей, законов, традиций называются *нормалами*. Именно нормали позволяют сопоставить факты, выявить и разрешить противоречия, сформировать новые факты-следствия, интегрировать данные в общую картину. Лишь в редких случаях нормаль может быть формализована и представлена в форме, допускающей ее автоматическую обработку. Большинство из них задается неформальными текстами или даже отсылкой на некоторое неформальное знание.

Применительно к процедуре очистки данных механизм нормалей позволяет осуществлять контроль корректности как при вводе информации, так и при осуществлении процедуры поиска противоречий [16]. В частности, нормали позволяют фиксировать каноническую форму имени и увязывать ее с приведенной в историческом источнике, что облегчает решение проблемы вариативности написания имен.

3.2 Связывание записей

Связывание записей (record linkage) — классическая проблема в исторической информатике, целью которой ставится определение факта, что две записи в разных исторических источниках относятся к одному объекту. Задача автоматизированного связывания записей для информации, полученной из номинативных исторических источников, довольно успешно решается в рамках метода восстановления истории семей (ВИС), используемого в исторической демографии [12]. В задачах связывания записей используются вероятностные алгоритмы. Сравнение отдельных атрибутов проводится с использованием фонетического кодирования для сравнения слов на основе их фонетического сходства, а также с применением методов машинного обучения. Для атрибутов задаются весовые коэффициенты, определяющие влияние на итоговую оценку вероятности того, что оцениваемые записи относятся к одному и тому же объекту. Пары записей с вероятностями выше некоторого порога считаются связанными, а пары с вероятностями ниже другого порога считаются несвязанными. Пары, которые попадают между этими двумя пороговыми значениями, считаются кандидатами на связывание и обрабатываются отдельно, для чего в технологии ПККИ предусмотрена надстройка над алгоритмами ВИС. Эта надстройка осуществляет оценку соответствия с опорой на два алгоритма [17]:

- (1) первый алгоритм помимо сравнения атрибутов отдельных записей опирается на оценку сходства структур родственных связей, образующих направленный граф с именованными связями и вершинами, соответствующими упоминаемым в документе персонам;
- (2) второй алгоритм для принятия решения о связывании записей опирается на теорию нечетких множеств и использование мер сходства.

Применение описанного механизма позволяет устранять несогласованность в данных посредством установления идентичности атрибутов, в том числе и имеющих отличия в написании, двух записей через их связывание (отнесение к одному реальному объекту).

3.3 Геоинформационные технологии

В технологии ПККИ предусмотрены механизмы поддержки микроанализа пространственных процессов, представленных в виде индивидуальных траекторий: геохронологический трекинг. Построение геохронологического трека исторической личности (индивида) или исторического объекта на основании геопространственной интерпретации его биографической информации есть интеграция хронологических и географических данных в виде графа, соединяющего географические точки нахождения исторической личности (группы и пр.) с цветоградиентной привязкой к тем или иным параметрам этого индивида или исторических событий. При этом вершины такого графа имеют строгую историко-географическую привязку, а дуги носят условно-логический характер [18].

Для этого в технологию ПКИИ интегрированы механизмы поддержки геоинформационной системы (ГИС). Для установления связей между семантической сетью технологии и географической БД ГИС в объектную модель технологии введено понятие ГИС-связки, состоящей из позиции, темпоральных границ связки и ссылки на источник информации. Помимо решения задач, возникающих в ходе конкретно-исторического исследования, ГИС способна решать проблему противоречивости и несогласованности пространственно-временных данных [19].

4 Схема очистки данных

На основании приведенного выше анализа встроенных в технологию ПКИИ средств можно выделить три этапа очистки данных:

- (1) при вводе данных;
- (2) при исправлении данных;
- (3) при формировании выборки для анализа.

К первому этапу относятся следующие средства:

- нормали (контроль корректности данных при вводе);
- наследование формата данных из сети классов (приведение данных к фиксированному формату при вводе).

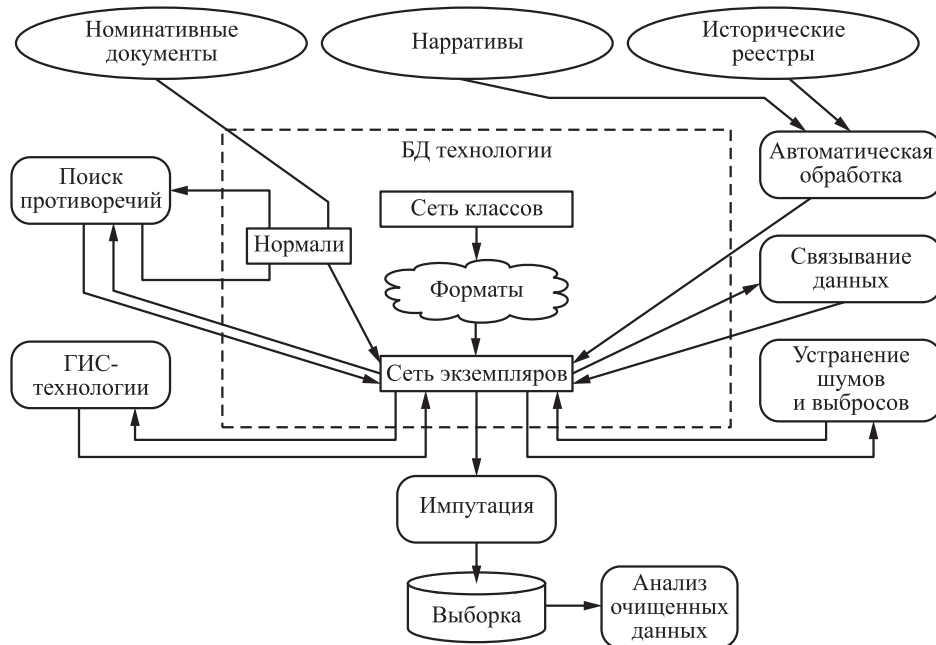


Схема очистки данных

Ко второму этапу относятся:

- кластеризация (устранение шумов и выбросов в данных);
- нормализация (поиск противоречий в данных);
- связывание записей (устранение несогласованности в данных);
- ГИС-технологии (устранение пространственно-временных ошибок и несогласованностей в данных).

К третьему этапу относится импутация (восстановление пропущенных данных).

Общая схема очистки данных в технологии ПКИИ приведена на рисунке.

5 Выводы

Как показано выше, применение алгоритмов машинного обучения в рамках конкретно-исторического исследования может значительно расширить арсенал исследователя. Поэтому их интеграция в технологию ПКИИ представляется перспективной задачей. Для успешного применения этих алгоритмов данные должны быть надлежащим образом подготовлены, и один из важнейших этапов такой подготовки — очистка данных. Как показано в статье, технология ПКИИ уже имеет в своем составе механизмы, достаточные для проведения очистки данных, что позволяет провести модернизацию технологии в указанном направлении. Актуальность дальнейшего развития технологии ПКИИ, ориентированной на широкий круг не относящихся к профессиональным историкам и биографам пользователей, вытекает из все возрастающего общественного интереса к частной, семейной истории.

Литература

1. *Грибач С. В.* Исследование семейных кризисов посредством психолингвистического эксперимента // Сборники конференций НИЦ Социосфера, 2010. № 6. С. 45–54.
2. *Адамович И. М., Волков О. И.* Технология распределенного автоматизированного анализа исторических текстов // Системы и средства информатики, 2016. Т. 26. № 3. С. 148–161. doi: 10.14357/08696527160311.
3. *Адамович И. М., Волков О. И.* Единая технология поддержки конкретно-исторических исследований // Системы и средства информатики, 2019. Т. 29. № 1. С. 194–205. doi: 10.14357/08696527190116.
4. *Адамович И. М., Волков О. И.* Принципы организации данных для технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2019. Т. 29. № 2. С. 161–171. doi: 10.14357/08696527190214.
5. *Фомина Е. Е.* Обзор методов и программного обеспечения для восстановления пропущенных значений в массивах социологических данных // Гуманитарный вестник, 2019. № 4(78). С. 1–12. doi: 10.18698/2306-8477-2019-4-611.

6. Гагарина Д. А., Корниенко С. И., Поврозник Н. Г. Информационные системы в цифровой среде исторической науки // История, 2017. № 7(51). <https://www.researchgate.net/publication/330242491>.
7. Воронина В. В., Михеев А. В., Ярушкина Н. Г., Святков К. В. Теория и практика машинного обучения. — Ульяновск: УлГТУ, 2017. 290 с.
8. Rahm E., Do H. H. Data cleaning: Problems and current approaches // IEEE Data Eng. Bull., 2000. Vol. 23. Iss. 4. P. 3–13.
9. Kim W., Choi B.-J., Hong E.-K., Kim S.-K., Lee D. A taxonomy of dirty data // Data Min. Knowl. Disc., 2003. Vol. 7. P. 81–99.
10. Osborne J. W. Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. — Newbury Park, CA, USA: SAGE Publications Inc., 2012. 275 p.
11. Адамович И. М., Волков О. И. Иерархическая форма представления биографического факта // Системы и средства информатики, 2016. Т. 26. № 2. С. 108–122. doi: 10.14357/08696527160207.
12. Адамович И. М., Волков О. И. Структурный подход к связыванию записей в технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2022. Т. 32. № 1. С. 94–103. doi: 10.14357/08696527220109.
13. Адамович И. М., Волков О. И. Механизм формирования гипотез в технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2023. Т. 33. № 1. С. 135–145. doi: 10.14357/08696527230113.
14. Ковалёв С. П. Использование алгоритма кластеризации DBSCAN для фильтрации выбросов в данных // Компьютерные системы и сети: 55-я юбилейная научная конференция аспирантов, магистрантов и студентов. — Минск: БГУИР, 2019. С. 198–200.
15. Адамович И. М., Волков О. И. Алгоритмы кластеризации для технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2022. Т. 32. № 4. С. 112–123. doi: 10.14357/08696527220411.
16. Адамович И. М., Волков О. И. Автоматизированный поиск противоречий в конкретно-исторической информации // Системы и средства информатики, 2020. Т. 30. № 3. С. 145–153. doi: 10.14357/08696527200313.
17. Адамович И. М., Волков О. И. Подход к связыванию записей в технологии поддержки конкретно-исторических исследований, основанный на нечетких множествах // Системы и средства информатики, 2022. Т. 32. № 2. С. 137–145. doi: 10.14357/08696527220213.
18. Адамович И. М., Волков О. И. Использование геоинформационных систем в технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2021. Т. 31. № 3. С. 158–169. doi: 10.14357/08696527210314.
19. Ерёмченко Е. Н., Тикунов В. С., Сун Чи-Гон. Противоречивость и несогласованность пространственно-временных данных: возможность решения проблемы в геоинформационной среде // Геодезия и картография, 2013. № 4. С. 41–47.

Поступила в редакцию 02.05.23

DATA CLEANSING IN THE TECHNOLOGY OF CONCRETE HISTORICAL INVESTIGATION SUPPORT

I. M. Adamovich and O. I. Volkov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The article continues the series of works devoted to the technology of concrete historical research supporting. The technology is based on the principles of co-creation and crowdsourcing and is designed for a wide range of users which are not professional historians and biographers. The expediency of expanding the list of concrete historical research tasks solved within the framework of the described technology using machine learning methods is shown. The special importance of data preparation is noted due to the fragmentation and inconsistency of concrete historical information. This article is devoted to the specifics of concrete historical data cleansing and the analysis of the possibility of using mechanisms and algorithms already integrated into the technology for this purpose. The main directions in which data cleansing is carried out are listed. Suitable tools already included in the technology have been identified for each direction. Particular attention is paid to tools for eliminating inconsistencies. The stages of data cleansing are listed and the scheme of interaction of all mechanisms and algorithms described in the article is given.

Keywords: concrete historical investigation; distributed technology; machine learning; data cleansing; data inconsistency

DOI: 10.14357/08696527230313

EDN: DDSTXY

References

1. Gribach, S. V. 2010. Issledovanie semeynykh krizisov posredstvom psikholingvisticheskogo eksperimenta [The study of family crises through a psycholinguistic experiment]. *Sborniki konferentsiy NITs Sotsiosfera* [Conference Proceedings NIC Sociosfera] 6: 45–54.
2. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelenno avtomatizirovannogo analiza istoricheskikh tekstov [The distributed automated technology of historical texts analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):148–161. doi: 10.14357/08696527160311.
3. Adamovich, I. M., and O. I. Volkov. 2019. Edinaya tekhnologiya podderzhki konkretno-istoricheskikh issledovaniy [Unified technology of concrete historical research support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):194–205. doi: 10.14357/08696527190116.
4. Adamovich, I. M., and O. I. Volkov. 2019. Printsipy organizatsii dannykh dlya tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [The principles of data organization for the technology of concrete historical research support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):161–171. doi: 10.14357/08696527190214.

5. Fomina, E. E. 2019. Obzor metodov i programmogo obespecheniya dlya vosstanovleniya propushchennykh znacheniy v massivakh sotsiologicheskikh dannykh [Review of software and methods for recovering missing values in sociological data sets]. *Gumanitarnyy vestnik* [Humanities Bulletin] 4(78):1–12. doi: 10.18698/2306-8477-2019-4-611.
6. Gagarina, D. A., S. I. Kornienko, and N. G. Povroznic. 2017. Informatsionnye sistemy v tsifrovoy srede istoricheskoy nauki [Information systems in the digital environment of historical studies]. *Istoriya* [History]. 7(51). Available at: <https://history.jes.su/s207987840001638-0-1/> (accessed August 16, 2023).
7. Voronina, V. V., A. V. Mikheev, N. G. Yarushkina, and K. V. Svyatov. 2017. *Teoriya i praktika mashinnogo obucheniya* [Theory and practice of machine learning]. Ulyanovsk: UIGTU. 290 p.
8. Rahm, E., and H. H. Do. 2000. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.* 23(4):3–13.
9. Kim, W., B.-J. Choi, E.-K. Hong, S.-K. Kim, and D. Lee. 2003. A taxonomy of dirty data. *Data Min. Knowl. Disc.* 7(1):81–99. doi: 10.1023/A:1021564703268.
10. Osborne, J. W. 2012. *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Newbury Park, CA: SAGE Publications Inc. 275 p.
11. Adamovich, I. M., and O. I. Volkov. 2016. Ierarkhicheskaya forma predstavleniya biograficheskogo fakta [Hierarchical format of a biographical fact]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(2):108–122. doi: 10.14357/08696527160207.
12. Adamovich, I. M., and O. I. Volkov. 2022. Strukturnyy podkhod k svyazyvaniyu zapisey v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [Structural approach to record linking in technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):94–103. doi: 10.14357/08696527220109.
13. Adamovich, I. M., and O. I. Volkov. 2023. Mekhanizm formirovaniya gipotez v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [Hypothesis formation mechanism in the technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(1):135–145. doi: 10.14357/08696527230113.
14. Kovalev, S. P. 2019. Ispol'zovanie algoritma klasterizatsii DBSCAN dlya fil'tratsii vybrosov v dannykh [Using the DBSCAN clustering algorithm for filtering outliers in data]. *Komp'yuternye sistemy i seti: 55-ya Yubileynaya nauchnaya konferentsiya aspirantov, magistrantov i studentov* [55th Anniversary Scientific Conference of Graduate Students, Undergraduates and Students “Computer Systems and Networks” Proceedings]. Minsk: BSUIR. 198–200.
15. Adamovich, I. M., and O. I. Volkov. 2022. Algoritmy klasterizatsii dlya tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [Clustering algorithms for technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(4):112–123. doi: 10.14357/08696527220411.
16. Adamovich, I. M., and O. I. Volkov. 2020. Avtomatizirovannyi poisk protivorechiy v konkretno-istoricheskoy informatsii [Automated search for contradictions in concrete-historical information]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(3):145–153. doi: 10.14357/08696527200313.

17. Adamovich, I. M., and O. I. Volkov. 2022. Podkhod k svyazyvaniyu zapisey v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy, osnovanny na nechetkikh mnozhestvakh [Approach to record linking in technology of concrete historical investigation support based on fuzzy sets]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):137–145. doi: 10.14357/08696527220213.
18. Adamovich, I. M., and O. I. Volkov. 2021. Ispol'zovanie geoinformatsionnykh sistem v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [The use of geographic information systems in technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(3):158–169. doi: 10.14357/08696527210314.
19. Eremchenko, E. N., V. S. Tikunov, and Chi-Gon Sun. 2013. Protivorechivost' i nesoglasovannost' prostranstvenno-vremennykh dannykh: vozmozhnost' resheniya problemy v geoinformatsionnoy srede [Conflicting and inconsistent spatio-temporal data: Problem solving ability in a geographic information environment]. *Geodeziya i kartografiya* [Geodesy and Cartography] 4:41–47.

Received May 2, 2023

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; volkov@amsd.com

СОЦИАЛЬНАЯ ЭФФЕКТИВНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

*К. К. Колин*¹

Аннотация: Рассмотрена проблема оценки эффективности информационных технологий и создаваемых на их основе информационных систем различного назначения. Показано, что их использование позволяет существенно повысить качество жизни и поэтому стало неотъемлемой частью жизнедеятельности современного общества. Кроме того, их применение необходимо для обеспечения общественной, национальной и глобальной безопасности. Приведен состав наиболее значимых в социальном аспекте перспективных информационных технологий и систем для различных сфер жизнедеятельности общества. Предложен концептуальный подход к созданию научной методологии оценки социальной эффективности информационных технологий и систем. Показано, что универсальным критерием для такой оценки может служить экономия социального времени, которая достигается в результате их практического использования. Рассмотрены перспективные направления развития информационных технологий.

Ключевые слова: информационные системы; информационные технологии; качество жизни; национальная безопасность; социальная эффективность

DOI: 10.14357/08696527230314

EDN: BHTQOX

1 Введение

Исследования показывают, что XXI в. стал началом новой эпохи развития цивилизации, в которой ключевыми факторами выступают информационные технологии. Они широко используются в самых разных сферах жизнедеятельности общества, выступая в них в качестве катализаторов развития, а также интеллектуальных компонентов других технологий — технических, энергетических, биологических и социальных.

В этих условиях оценка социальной эффективности информационных технологий и создаваемых на их основе информационных систем становится актуальной и стратегически важной проблемой. Ее решение необходимо для оптимизации государственной политики и деятельности структур бизнеса в области информационного развития общества и определения приоритетных направлений этого развития.

Анализ состояния исследований в этой области показал, что создание научно обоснованной методологии оценки социальной эффективности информационных

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, kolinkk@mail.ru

технологий и систем находится еще в начальной стадии. Содержание этой проблемы требует для своего решения системного подхода. Попытка рассмотреть концептуальные положения такого подхода и ставится основной целью настоящей работы.

2 Современное состояние проблемы

Одна из первых попыток анализа проблемы социальной эффективности информационных технологий была сделана в Институте проблем информатики РАН в 1995 г. В работе [1] показано, что информационные технологии служат катализаторами социально-экономического развития общества, так как они становятся не только основными средствами для использования информационных ресурсов общества, но также и существенным образом повышают эффективность других видов технологий, выступая в них в качестве интеллектуального ядра.

В этой работе предложено различать следующие классы технологий:

- **производственные технологии**, предназначенные для оптимизации процессов в сфере материального производства товаров и услуг и их общественного распределения;
- **информационные технологии**, предназначенные для рациональной организации процессов в информационной сфере общества, включая науку, культуру, образование, средства массовой информации и информационные коммуникации;
- **социальные технологии**, ориентированные на рациональную организацию социальных процессов в экономике, политике, финансовой сфере, образовании, здравоохранении, военном деле и т. п.;

Была также показана необходимость формирования специальной *науки о технологиях* и определены ее задачи и направления исследований. Эта проблема была затем детально рассмотрена в работе [2].

Основа концептуального подхода к изучению этой проблемы состоит в использовании понятия *социального времени*, которое было предложено академиком В. Г. Афанасьевым в качестве универсальной меры затрат общественного труда на реализацию любых социально-экономических процессов. Это понятие в дальнейшем использовал П. Г. Кузнецов [3].

Исследования показали, что понятие социального времени целесообразно использовать при оценке социальной эффективности любых видов технологий, включая информационные [4].

3 Классификация информационных технологий

В настоящее время классификация информационных технологий, применяемых в различных сферах социальной практики, осуществляется в основном

по признакам их практического использования, т. е. из чисто прагматических соображений. Однако анализ технологий с точки зрения их социальной эффективности требует иного подхода. В его основе должен лежать характер использования этих технологий в процессе реализации информационно-технологических процессов.

С этой точки зрения целесообразно выделить следующие два основных класса информационных технологий:

- (1) *базовые информационные технологии;*
- (2) *прикладные информационные технологии.*

Базовые информационные технологии представляют собой наиболее эффективные способы организации типовых, часто повторяющихся информационных процессов, связанных с преобразованием и использованием тех или иных видов информационных ресурсов. Примерами таких технологий могут быть технологии кодирования информации, ее хранения и отображения в различных формах, а также технологии передачи информации по каналам связи.

Характерный признак базовых технологий — они не предназначены для непосредственной реализации тех или иных конкретных информационных технологических процессов, а служат лишь теми базовыми их компонентами, на основе которых и проектируются затем прикладные информационные технологии.

Главная цель базовых информационных технологий заключается в достижении *максимальной эффективности* в реализации некоторого фрагмента более сложного информационно-технологического процесса на основе использования последних достижений науки и технологий, а также передового практического опыта. Именно поэтому для базовых технологий и разрабатываются национальные и международные стандарты и протоколы, использование которых гарантирует высокое качество реализации информационно-технологических процессов.

Прикладные информационные технологии. Основной задачей здесь ставится рациональная организация некоторого конкретного информационно-технологического процесса. Осуществляется это путем адаптации нескольких базовых технологий, позволяющих наилучшим образом реализовать отдельные фрагменты этого процесса. Поэтому основными научными проблемами в области исследования прикладных информационных технологий можно считать:

- (1) разработку методов анализа, синтеза и оптимизации прикладных информационных технологий;
- (2) создание теории проектирования информационных технологий различного вида и практического назначения;
- (3) создание методологии сравнительной количественной оценки различных вариантов построения технологий, а также их эффективности;
- (4) разработку требований к средствам реализации прикладных информационных технологий.

Одним из примеров прикладной информационной технологии может служить технология ввода в память компьютера речевой информации. С технологической точки зрения информационный процесс здесь разделяется на несколько последовательных этапов, на каждом из которых используется своя базовая информационная технология.

Основной принцип формирования прикладной информационной технологии — адаптация заранее отработанных базовых технологий для реализации данного информационно-технологического процесса. Это дает большую экономию времени при разработке прикладных информационных технологий и гарантирует их высокую эффективность.

4 Критерии эффективности информационных технологий

Для оценки эффективности различных вариантов информационных технологий нужно уметь правильно выбирать критерии их эффективности. Такими критериями могут быть:

- (1) **функциональные критерии**, которые характеризуют степень достижения желаемых характеристик информационно-технологического процесса при использовании данной технологии;
- (2) **ресурсные критерии**, которые характеризуют объем и качество различного вида ресурсов, необходимых для реализации данной технологии. Такими ресурсами могут быть:
 - *материальные ресурсы* (сырье, материалы, технологическое оборудование);
 - *энергетические ресурсы* (затраты энергии на реализацию информационно-технологического процесса при данной технологии);
 - *людские ресурсы* (численность и уровень подготовки персонала, необходимого для реализации данной технологии);
 - *временные ресурсы* (количество времени, необходимого для реализации процесса при использовании данной технологии);
 - *информационные ресурсы* (данные и знания, необходимые для успешной реализации информационно-технологического процесса).

В настоящее время основными видами ресурсов в производственной сфере выступают материальные и энергетические ресурсы. Именно поэтому наибольшее внимание в отраслях производства промышленной продукции уделяется так называемым материалосберегающим и энергосберегающим производственным технологиям. Что же касается информационных технологий, то здесь имеется своя достаточно существенная специфика.

Так, энергетические ресурсы для информационных технологий, как правило, имеют второстепенное значение. Ведь информационные процессы обладают

сравнительно низкой энергоемкостью по сравнению с силовыми процессами, которые реализуются в механических и энергетических технологиях.

Отметим также, что информационные технологии стали основным средством формирования и использования *информационных ресурсов* общества. Однако их особенность заключается в том, что для своего функционирования они сами нуждаются в использовании информационных ресурсов. Эти ресурсы в виде баз данных и знаний могут заранее вводиться в память информационной системы, а также поступать в нее извне в процессе реализации того или иного информационно-технологического процесса.

Характерным примером здесь служат *экспертные системы*. Эти технологии, как правило, используют уже накопленный опыт в организации того или иного информационного процесса. При этом достигается возможность существенным образом снизить уровень требований к профессиональной квалификации пользователей экспертной системы, что дает значительный экономический и социальный эффект.

Этот пример показывает, что информационные технологии позволяют не только формировать знания, но также и их экономно использовать. Другими словами, они обладают свойствами *информационно сберегающих технологий*. Никакие другие технологии такими свойствами не обладают.

Общий критерий эффективности информационных технологий. Ресурсные критерии эффективности позволяют сравнивать между собой различные виды технологий. Кроме того, они дают возможность количественно оценивать получаемый в результате применения этих технологий эффект с точки зрения их социальной полезности — в плане экономии различных видов ресурсов общества. Именно поэтому наиболее распространенными критериями для оценки производственных технологий стали *энергетические критерии*. Ведь затраты энергии в общественно полезном производстве служат одним из важнейших показателей уровня технологического развития современного общества.

Однако наиболее общим показателем технологии любого вида (производственной, социальной или же информационной) следует признать **экономии социального времени**. Этот критерий предложен В. Г. Афанасьевым и П. Г. Кузнецовым в качестве одной из наиболее общих мер развития общества [5]. Представляется, что он пригоден для количественной оценки эффективности самых разных видов технологий. Ведь хорошо известно, что любая экономия в конечном итоге может быть сведена к экономии времени.

Мало того, по мнению П. Г. Кузнецова, именно *бюджет социального времени* и есть главный ресурс для жизнеобеспечения и развития общества [3]. Согласно этому критерию, более полезны для общества те технологии, которые позволяют сэкономить больше социального времени, высвобождая его для других целей, в том числе для целей развития самого общества.

Изложенный подход коренным образом изменяет традиционную точку зрения на эффективность тех или иных видов информационных технологий, которые

сегодня оцениваются, как правило, лишь по функциональным критериям. Так, с точки зрения экономии социального времени для общества очень эффективным оказывается использование информационных технологий в сфере *массового обслуживания населения* (на предприятиях торговли, общественного питания, в сберегательных банках, билетных кассах и т. п.).

Об этом убедительно свидетельствует практика создания в нашей стране многофункциональных центров оказания населению государственных услуг. Масштабы экономии социального времени, которое ранее затрачивалась на получение этих услуг, огромны. И это очень важное направление повышения качества жизни граждан России.

Использование экономии социального времени в качестве общего критерия эффективности информационных технологий сегодня еще не обеспечено необходимыми методическими разработками. Однако данный подход представляется исключительно перспективным. Он изменяет мировоззрение общества, его отношение к социальной роли информационных технологий в различных сферах жизнедеятельности.

5 Социально значимые информационные технологии и системы

В настоящее время происходит глубокая трансформация многих сфер жизнедеятельности общества, обусловленная широким распространением в них новых информационных технологий, систем и телекоммуникаций. Характер этих перемен в сжатой форме показан в виде таблицы, где приведены те современные и перспективные информационные технологии и системы, которые уже сегодня оказывают сильное воздействие на общество и, безусловно, сохранят это влияние в будущем.

6 Перспективы развития информационных технологий

В последние годы наиболее динамичным и стратегически социально значимым стало развитие тех *информационных технологий*, которые служат катализатором других процессов развития современного общества. Результатом этого стал массовый переход к цифровым технологиям и становление цифровой экономики [6].

На этой основе в экономически развитых странах уже начался переход к новому, шестому, технологическому укладу. Его отличительной особенностью будет *интеграция различных видов технологий*, а также их все более высокая интеллектуализация на основе использования *базовых технологий искусственного интеллекта*. При этом ожидается существенное повышение эффективности интегрированных технологий, а также их проникновение в новые области жизнедеятельности общества, которые в результате этого претерпят революционные изменения [7].

Например, интеграция технологий использования солнечной энергии с интеллектуальными информационными технологиями позволяет создать новую сис-

Состав социально значимых современных и перспективных информационных технологий и систем

Сферы деятельности	Социально значимые информационные технологии и системы
Общественная безопасность	Космические системы мониторинга опасных процессов. Системы телевизионного наблюдения в городах и на транспорте. Контроль опасных грузов в аэропортах, на вокзалах, на таможне. Системы оповещения населения об опасных ситуациях по радио, телевидению и мобильной связи. Системы идентификации личности. Охранная сигнализация объектов и помещений
Национальная оборона	Системы противовоздушной, противокосмической и противоракетной обороны. Информационная разведка и противодействие. Системы целеуказания и наведения. Управление беспилотными аппаратами и боеприпасами. Интеллектуальные технологии в системах управления национальной обороной. Ситуационные аналитические центры. Боевые роботы
Общественное производство и распределение	Промышленные роботы. Умные производства и фабрики. Цифровые двойники промышленных изделий. Системы электронной торговли и распространения продукции. Реклама
Экономика и финансы	Ситуационные центры управления экономикой. Электронные платежные системы. Информационные банковские системы
Транспорт	Космические навигационные системы. Системы управления беспилотным наземным, воздушным и водным транспортом
Социальная сфера	Информационные центры государственных услуг, их личные кабинеты в сети Интернет. Электронные медицинские карты. Технологии телемедицины. Медицинские роботы
Информационная сфера	Мобильная связь, телевидение и радиовещание. Социальные сети. Цифровые платформы. Компьютерный перевод текстов и речи. Биологические экраны. Квантовые и оптические коммуникации. Искусственный интеллект, нейронные сети
Наука и образование	Суперкомпьютеры. Квантовые компьютеры. Большие данные. Когнитивные технологии. Полиэкранные педагогические технологии. Семантический поиск информации. Банки знаний
Культура	Электронные библиотеки. Виртуальные музеи и галереи. Электронная культура. Голографическая анимация. Информационные технологии искусства и творчества

тему электроснабжения зданий и помещений, которая существенно сократит потребление углеводородного топлива и создаст условия для перехода к «горизонтальной структуре» промышленного производства товаров и услуг. С этой целью в Западной Европе реализуется специальная программа «зеленой энергетики», первый этап которой завершен в 2020 г. Эта программа стала первым этапом воплощения в Европе новой энергетической концепции *Третьей промышленной революции* [8].

Развитие «зеленой энергетики» дает и весьма ощутимый *социальный эффект*, создавая новые рабочие места в области новых энергетических и информационных технологий. Так, в Германии в этой области в 2007 г. было занято 250 тыс. специалистов — почти столько же, что и в углеводородной энергетике. Однако подлинная революция ожидается в результате *интеграции аддитивных биологических технологий с информационными технологиями* [9]. Она позволит создавать искусственные органы и имплантировать их людям в медицинской практике.

7 Сетевое информационно-технологическое общество

Признаки становления сетевого общества в последние годы проявляются все более отчетливо не только в экономически развитых, но также и во многих развивающихся странах, например в Китае, Индии, Малайзии. По имеющимся прогнозам, уже к середине XXI в. сетевые структуры будут доминировать во многих сферах — в науке, образовании, культуре, здравоохранении, социальных коммуникациях. При этом вполне вероятно, что доступ к сетевым технологиям, а также многие информационные услуги будут предоставляться бесплатно. Поэтому представления о качестве жизни, а также о личном и национальном богатстве существенно изменятся. Уже сегодня эти технологии и услуги стали атрибутами нашей повседневной жизни и деятельности, важной частью современной культуры общества. Исследования показывают, что в дальнейшем их значимость будет только возрастать.

Специалисты прогнозируют взрывообразный характер развития информационной сферы общества, который будет обусловлен массовым использованием «интернета вещей» в производственных структурах, а также в бытовой сфере. Ожидается, что уже в 2025 г. к сети Интернет будут подключены более 1 трлн таких устройств, а трафик обмена данными между ними превысит потоки информации между людьми.

Еще одним стратегически важным направлением развития информационных технологий стала их реализация в *ситуационных центрах*, которые в последние годы получают все более широкое применение в сфере государственного управления на национальном, региональном, муниципальном и ведомственном уровнях [10].

8 Заключение

Уровень развития технологий сегодня характеризует не только развитие экономики той или иной страны, но также и ее место в мировом сообществе. В ближайшие годы можно ожидать создание новых информационных технологий, эффективность которых будет превышать современный уровень на порядки.

Их появление будет означать «прорыв» на качественно новый уровень технологического развития [11].

Именно поэтому сегодня необходимо, чтобы была сформирована *новая научная дисциплина — информационная технология*. Она должна стать научной базой для развития *информационной цивилизации*. Формирование этого направления позволит повысить эффективность использования информационных ресурсов общества, а также существенно сократить затраты социального времени на реализацию многих процессов его жизнеобеспечения. Предмет и задачи этой научной дисциплины были определены в Институте проблем информатики РАН еще в 2001 г. [2].

Интеграция сетевых технологий и технологий искусственного интеллекта изменит весь облик информационной сферы общества. Такие понятия, как «умный автомобиль», «умный дом», «умный завод», «умный город» и т. п., прочно войдут в нашу жизнь и станут привычными. Роботизация облегчит труд миллионов людей и сделает его более безопасным, а новые технологии обработки текстов навсегда избавят человечество от рутинной канцелярской работы.

Однако информатизация общества влечет за собой не только новые социальные блага и возможности, но также новые вызовы и угрозы, признаки которых проявляют себя в последние годы [12–14]. И эта проблематика должна постоянно находиться в центре внимания ученых, педагогов и государственных деятелей современного общества.

Литература

1. *Колин К. К.* Информационные технологии — катализатор развития современного общества // Информационные технологии, 1995. № 0. С. 2–8.
2. *Колин К. К.* Информационная технология как научная дисциплина // Информационные технологии, 2001. № 2. С. 2–10.
3. *Кузнецов П. Г.* Бюджет социального времени // По ту сторону отчуждения. — М.: МГУ им. М. В. Ломоносова, 1992. С. 227–251.
4. *Колин К. К.* Информационные проблемы социально-экономического развития общества. — М.: Союз, 1995. 72 с.
5. *Гвардейцев М. И., Кузнецов П. Г., Розенберг В. Я.* Математическое обеспечение управления. Меры развития общества. — М.: Радио и связь, 1996. 176 с.
6. *Колин К. К.* Физическое изменение цифровой экономики и социальная эффективность новых технологий // Информационные ресурсы России, 2018. № 5. С. 2–9.
7. *Колин К. К.* Качество жизни в информационном обществе // Человек и труд, 2010. № 1. С. 39–43.
8. *Рифкин Дж.* Третья промышленная революция. Как горизонтальные взаимодействия меняют экономику, энергетику и мир в целом / Пер. с англ. — М.: Альпина нон-фикшн, 2015. 410 с. (*Rifkin J.* The third industrial revolution: How lateral power is transforming energy, the economy, and the world. — New York, NY, USA: Palgrave Macmillan, 2011. 304 p.)

9. Шваб К. Четвертая промышленная революция / Пер. с англ. — М.: Эксмо, 2017. 208 с. (*Schwab K. The fourth industrial revolution. — New York, NY, USA: Crown Publishing Group, 2016. 192 p.*)
10. Стратегическое целеполагание в ситуационных центрах развития / Под ред. В. Е. Лепского и А. Н. Райкова. — М.: Когито-Центр, 2018. 320 с.
11. Яковец Ю. В., Колин К. К. Стратегия научно-технологического прорыва России. — М.: Сокол, 2015. Вып. 7. 51 с.
12. Колин К. К. Проблемы информационной цивилизации: виртуализация общества // Библиоковедение, 2002. № 3. С. 48–57.
13. Колин К. К. Информационная безопасность как гуманитарная проблема // Открытое образование, 2006. № 1. С. 86–93.
14. Колин К. К. Интеллектуальная безопасность — новая глобальная проблема XXI века // Стратегические приоритеты, 2019. № 3-4. С. 99–111.

Поступила в редакцию 07.04.23

SOCIAL EFFICIENCY OF INFORMATION TECHNOLOGIES

K. K. Kolin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The problem of evaluating the effectiveness of information technologies and information systems for various purposes created on their basis is considered. It is shown that their use can significantly improve the quality of life and, therefore, has become an integral part of the life of modern society. In addition, their use is necessary to ensure public, national, and global security. The composition of the most significant in the social aspect promising information technologies and systems for various spheres of society is given. A conceptual approach to the creation of a scientific methodology for assessing the social effectiveness of information technologies and systems is proposed. It is shown that the saving of social time which is achieved as a result of their practical use can serve as a universal criterion for such an assessment. Perspective directions of development of information technologies are considered.

Keywords: information systems; information technologies; quality of life; national security; social efficiency

DOI: 10.14357/08696527230314

EDN: BHTQOX

References

1. Kolin, K. K. 1995. Informatsionnye tekhnologii — katalizator razvitiya sovremennogo obshchestva [Information technology — a catalyst for development of modern society]. *Informatsionnye tekhnologii* [Information Technologies] 0:2–8.
2. Kolin, K. K. 2001. Informatsionnaya tekhnologiya kak nauchnaya distsiplina [Information technology as a scientific discipline]. *Informatsionnye tekhnologii* [Information Technologies] 2:2–10.

3. Kuznetsov, P. G. 1992. Byudzhet sotsial'nogo vremeni [The social time budget]. *Po tu storonu otchuzhdeniya* [On the other side of alienation]. Moscow: Moscow State University. 227–251.
4. Kolin, K. K. 1995. *Informatsionnye problemy sotsial'no-ekonomicheskogo razvitiya obshchestva* [Information problems of socio-economic development of society]. Moscow: Soyuz. 72 p.
5. Gvardeytshev, M. I., P. G. Kuznetsov, and V. Ya. Rozenberg. 1996. *Matematicheskoe obespechenie upravleniya. Mery razvitiya obshchestva* [Mathematical support of control. Measures of society development]. Moscow: Radio i svyaz'. 176 p.
6. Kolin, K. K. 2018. Fizicheskoe izmenenie tsifrovoy ekonomiki i sotsial'naya effektivnost' novykh tekhnologiy [Physical dimension of digital economy and social efficiency of new technologies]. *Informatsionnye resursy Rossii* [Information resources of Russia] 5:2–9.
7. Kolin, K. K. 2010. Kachestvo zhizni v informatsionnom obshchestve [The quality of life in the information society]. *Chelovek i trud* [Person and Labour] 1:39–43.
8. Rifkin, J. 2011. *The third industrial revolution: How lateral power is transforming energy, the economy, and the world*. New York, NY: Palgrave Macmillan. 304 p.
9. Schwab, K. 2016. *The fourth industrial revolution*. New York, NY: Crown Publishing Group. 192 p.
10. Lepskiy, V. E., and A. N. Raykov, eds. 2018. *Strategicheskoe tselepolaganiye v situatsionnykh tseentrakh razvitiya* [Strategic goal-setting in situational development centers]. Moscow: Cogito-Centre. 320 p.
11. Yakovets, Yu. V., and K. K. Kolin. 2015. *Strategiya nauchno-tekhnologicheskogo progressa Rossii* [Strategy of scientific and technological breakthrough of Russia]. Moscow: Sokol. No. 7. 51 p.
12. Kolin, K. K. 2002. Problemy informatsionnoy tsivilizatsii: virtualizatsiya obshchestva [Problems of information civilization: Virtualization of society]. *Bibliotekovedenie* [Russian J. Library Science] 3:48–57.
13. Kolin, K. K. 2006. Informatsionnaya bezopasnost' kak gumanitarnaya problema [Information security as a humanitarian issue]. *Otkrytoe obrazovanie* [Open Education] 1:86–93.
14. Kolin, K. K. 2019. Intellektual'naya bezopasnost' — novaya global'naya problema XXI veka [Intellectual security — a new global problem of the XXI century]. *Strategicheskie priority* [Strategic Priorities] 3-4:99–111.

Received April 7, 2023

Contributor

Kolin Konstantin K. (b. 1935) — Doctor of Science in technology, professor, honored scientist of RF, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; kolinkk@mail.ru

Абрамов Алексей Геннадьевич (р. 1976) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Санкт-Петербургского отделения Межведомственного суперкомпьютерного центра Российской академии наук — филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Барашов Егор Борисович (р. 1999) — инженер Института проблем управления им. В. А. Трапезникова Российской академии наук

Волков Олег Игоревич (р. 1964) — ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Егоркин Андрей Владимирович (р. 2001) — студент бакалавриата Национального исследовательского университета электронных технологий

Егорова Анна Юрьевна (р. 1991) — младший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; старший преподаватель Московского государственного лингвистического университета

Забжайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ильин Владимир Дмитриевич (р. 1937) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Колин Константин Константинович (р. 1935) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Лемтюжникова Дарья Владимировна (р. 1988) — кандидат физико-математических наук, доцент Московского авиационного института (национального исследовательского университета); старший научный сотрудник Института проблем управления им. В. А. Трапезникова Российской академии наук

Матвеев Иван Алексеевич (р. 1973) — доктор технических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Никишин Дмитрий Александрович (р. 1976) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Порхачёв Василий Александрович (р. 1967) — ведущий специалист Санкт-Петербургского отделения Межведомственного суперкомпьютерного центра Российской академии наук — филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»

Посыпкин Михаил Анатольевич (р. 1974) — доктор физико-математических наук, член-корреспондент РАН, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — доктор физико-математических наук, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

Романенко Варвара Олеговна (р. 1983) — инженер 1 категории Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сафонов Илья Владимирович (р. 1971) — кандидат технических наук, доцент Национального ядерного исследовательского университета «МИФИ»

Сучков Александр Павлович (р. 1954) — доктор технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шабанов Александр Петрович (р. 1949) — доктор технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положения глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или L^AT_EX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете:

<http://www.ipiran.ru/publications/collected/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:
 - название статьи;
 - Ф.И.О. авторов, на английском можно только имя и фамилию;
 - место работы, с указанием города и страны и электронного адреса каждого автора;
 - сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
 - аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
 - ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.
 - источники финансирования работы (ссылка на гранты, проекты, поддерживающие организации и т. п.).
9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала.
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povyshe-niya neftegazootdachi”* [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W.* Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York, NY: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktny tor [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatzionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Spособ orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: ssi@frccsc.ru (Стригина Светлана Николаевна)

<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . . ,” pass to the Founder and the Editorial Board of the Journal “Systems and Means of Informatics” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).”

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents

may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site:

<http://www.ipiran.ru/publication/collected/template.doc>

8. Articles should enclose data both in **Russian and English**:

- title;
- author's name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format (see site):
http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae.
- Indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences.
- Acknowledgments.

9. References. Russian references have to be presented both in English translation and in Latin transliteration (refer <http://www.translit.net/ru/bgn/>).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic

exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povyshe-niya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogenera-torov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovaryia informat-sionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
 - the journal title and author's name in the "Subject" field;
 - an article and additional materials have to be attached using the "attach" function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "System and Means of Informatics" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: ssi@frccsc.ru (to Svetlana Strigina)

http://www.ipiran.ru/english/journal_systems.asp

SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

Volume 33 No.3 Year 2023

Editor-in-Chief and Chair of Editorial Council

Academician I. A. Sokolov

IN THIS ISSUE:

TO THE 40TH ANNIVERSARY OF THE DECREES OF THE CENTRAL COMMITTEE OF THE CPSU AND THE COUNCIL OF MINISTERS OF THE USSR ON THE FURTHER DEVELOPMENT OF WORK IN THE FIELD OF COMPUTER TECHNOLOGY, INCLUDING AT THE USSR ACADEMY OF SCIENCES <i>V. N. Zakharov</i>	4
TIME SERIES MONOTONIC TREND ANALYSIS <i>M. P. Krivenko</i>	17
DISPATCHING IN NONOBSERVABLE PARALLEL QUEUES WITH FINITE CAPACITIES <i>M. G. Kononov and R. V. Razumchik</i>	29
PRINCIPLES OF DEVELOPMENT OF A SOFTWARE SUITE FOR THE NETWORK OPERATIONS CENTER OF THE NATIONAL RESEARCH COMPUTER NETWORK OF RUSSIA <i>A. G. Abramov and V. A. Porkhachev</i>	48
EFFICIENCY OF THE REDUCTION ALGORITHMS IN THE BIN PACKING PROBLEM <i>E. B. Barashov, A. V. Egorkin, D. V. Lemtyuzhnikova, and M. A. Posypkin</i>	61
RESEARCH METHODOLOGY FOR CREATING FUNCTIONS OF SCANNING AND PRINTING SYSTEMS <i>I. V. Safonov and I. A. Matveev</i>	76
PROBLEMS OF FORMATION OF A SYSTEM OF COORDINATED LOCAL GEOGRAPHICAL ONTOLOGIES <i>D. A. Nikishin</i>	85
SOME APPROACHES TO THE ANALYSIS OF FACTORS AFFECTING THE INFORMATION SECURITY OF ARTIFICIAL INTELLIGENCE SYSTEMS <i>A. A. Zatsarinny and A. P. Suchkov</i>	95