

# **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  
к.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  
к.ф.-м.н. Р. В. Разумчик  
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 № 1 Год 2023

## СОДЕРЖАНИЕ

Самосинхронный конвейер с переменным числом ступеней <b><i>И. А. Соколов, Ю. А. Степченков, Ю. Г. Дьяченко, Н. В. Морозов, Д. Ю. Дьяченко</i></b>	<b>4</b>
Корпусные данные при контрастивном изучении пунктуации <b><i>В. А. Нуриев, М. Г. Кружков</i></b>	<b>14</b>
Интеграционные возможности надкорпусных баз данных <b><i>А. А. Дурново, О. Ю. Инькова, В. А. Нуриев</i></b>	<b>24</b>
Использование базы данных моделей структурных переводческих трансформаций для извлечения многокомпонентных терминологических единиц <b><i>Ю. И. Бутенко</i></b>	<b>35</b>
Семантические интерпретации высоких нормальных форм отношений реляционной базы данных <b><i>В. А. Иванов, М. Ю. Конышев, С. В. Смирнов, О. В. Тараканов, В. О. Тараканова, С. В. Усовик</i></b>	<b>45</b>
Задача классификации в условиях искаженных причинно-следственных связей <b><i>А. А. Грушо, Н. А. Грушо, М. И. Забежайло, А. А. Зацаринный, Е. Е. Тимонина, С. Я. Шоргин</i></b>	<b>59</b>
Методы классификации пользователей СДО в модели построения их индивидуальной траектории обучения <b><i>Я. Г. Мартюшова, Т. А. Минеева, А. В. Наумов</i></b>	<b>68</b>
Эффективные вычисления при факторизации матричных данных с пропусками <b><i>М. П. Кривенко</i></b>	<b>78</b>
Об оценках устойчивости и их применении для некоторых моделей массового обслуживания <b><i>И. А. Ковалёв</i></b>	<b>90</b>

# СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 33 № 1 Год 2023

## СОДЕРЖАНИЕ

Программное определение сети в конвергентной и гиперконвергентной инфраструктурах <b>В. Б. Егоров</b>	<b>105</b>
Центры компетенции по искусственному интеллекту и Национальная технологическая инициатива <b>А. П. Сучков</b>	<b>114</b>
Теория S-символов: концептуальные основания <b>В. Д. Ильин</b>	<b>126</b>
Механизм формирования гипотез в технологии поддержки конкретно-исторических исследований <b>И. М. Адамович, О. И. Волков</b>	<b>135</b>
Базисные информационные технологии, информационная война и многомерный иерархический территориальный суверенитет: этапы глобально-космической коэволюции <b>С. Н. Гринченко</b>	<b>146</b>
Об авторах	<b>154</b>
Правила подготовки рукописей статей	<b>157</b>
Requirements for manuscripts	<b>161</b>

## САМОСИНХРОННЫЙ КОНВЕЙЕР С ПЕРЕМЕННЫМ ЧИСЛОМ СТУПЕНЕЙ\*

*И. А. Соколов<sup>1</sup>, Ю. А. Степченко<sup>2</sup>, Ю. Г. Дьяченко<sup>3</sup>, Н. В. Морозов<sup>4</sup>,  
Д. Ю. Дьяченко<sup>5</sup>*

**Аннотация:** Рассматривается проблема увеличения производительности самосинхронных схем (СС). Эффективным способом повышения производительности, как и в синхронных схемах, оказалось использование конвейера для реализации многоступенчатой обработки входных данных. Анализируются возможные варианты динамического сокращения числа активно работающих ступеней при выполнении некоторых условий, определяемых значением обрабатываемых данных или внешним сигналом. Приведенные оценки показывают, что эффективность использования опционально варьируемого числа ступеней конвейера зависит от числа маскируемых ступеней и вероятности появления события, разрешающего такое маскирование. В частности, замена двух последовательных ступеней конвейера одной параллельной ступенью становится целесообразной, если она реализуется не менее чем в 63% операций обработки данных, а обход двух и более ступеней уменьшает среднюю латентность конвейера, если он срабатывает не менее чем в 43% операций.

**Ключевые слова:** самосинхронная схема; конвейер; обход; мультиплексирование; латентность; производительность

**DOI:** 10.14357/08696527230101

### 1 Введение

Практические синхронные цифровые схемы проектируются в виде конвейера [1] для обеспечения максимального быстродействия. Конвейер хорошо ложится в русло глобальной синхронизации. Самосинхронные схемы [2] также целесообразно реализовывать в виде конвейера для повышения производительности.

---

\* Исследование выполнено в рамках гранта Российского научного фонда (проект 22-19-00237).

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ISokolov@ipiran.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, YStepchenkov@ipiran.ru

<sup>3</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, diaura@mail.ru

<sup>4</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, NMogozov@ipiran.ru

<sup>5</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, diaden87@gmail.com

Самосинхронные схемы не имеют глобальной синхронизации, так как их поведение основывается на запросно-ответном механизме взаимодействия составных частей общей схемы. Однако они содержат управляющие сигналы, регулирующие двухфазное взаимодействие СС-устройств. Управляющие сигналы отражают готовность ступеней СС-конвейера к переключению в следующую фазу работы и формируются сигналами, индицирующими части СС-схемы.

Двухфазная дисциплина работы и обязательное подтверждение индикаторной подсхемой окончания переключения СС-схемы в каждую фазу работы обеспечивают контроль корректности переключения схемы и повышают устойчивость СС-схемы к логическим сбоям [3]. Индикаторная подсхема реализуется в виде многоярусной пирамидальной схемы с одним выходом и с числом входов, равным числу контролируемых точек. По мере увеличения числа контролируемых точек число ярусов растет, что ведет к снижению быстродействия СС-схемы. Разбиение на ступени конвейера общей СС-схемы уменьшает сложность схемы, индицируемой в рамках одной ступени, и, соответственно, сложность и задержку ее индикаторной подсхемы. Длительность рабочего цикла одной ступени сокращается, и производительность СС-конвейера повышается.

Однако в многоразрядных вычислительных СС-устройствах проблема повышения производительности СС-конвейера по-прежнему актуальна. Данная статья рассматривает особенности взаимодействия ступеней СС-конвейера и предлагает принципы организации СС-конвейера, позволяющие ускорять прохождение данных по конвейеру в определенных случаях за счет опциональной регулировки числа активно работающих ступеней.

## 2 Типовой самосинхронный конвейер

На рис. 1 представлена четырехступенчатая схема типового СС-конвейера. Каждая ступень состоит из комбинационной части ( $C_i$ ) и выходного регистра ( $R_i$ ),

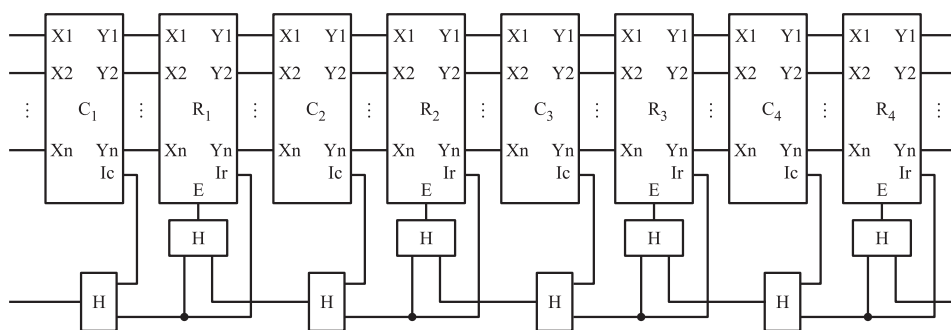


Рис. 1 Схема типового СС-конвейера

имеющих парафазные со спейсером входы ( $X_1, \dots, X_n$ ) и выходы ( $Y_1, \dots, Y_n$ ). Их индикаторные выходы  $I_s$  и  $I_r$  формируют сигнал управления  $E$  регистром предыдущей ступени с помощью гистерезисных триггеров ( $\Gamma$ -триггеров), обозначенных буквой «Н».

При выполнении определенных условий (например, при определенных значениях обрабатываемых данных) отдельные части ступеней конвейера или целые ступени могут выключаться из тракта обработки данных, если их участие не требуется: данные могут «обходить» их. Например, в устройстве умножения с накоплением (Fused Multiply-Add [4]) целесообразность вычисления произведения двух операндов, а также суммы и/или разности произведения и третьего операнда определяется их соотношением. Если третий операнд не имеет значения при заданной точности операции, то выполнять сложение/вычитание нет необходимости. Если произведение слишком мало, то и умножение делать не требуется. Оценка соотношения произведения и третьего операнда может быть получена путем сравнения их экспонент до начала вычисления произведения.

В синхронном конвейере мультиплексоры, обеспечивающие обход одной или нескольких ступеней, снижают частоту синхронизации, так как она рассчитывается для наихудшего случая, а мультиплексоры привносят дополнительную задержку.

В СС-конвейере нет необходимости учитывать наихудший случай. Его производительность характеризуется статистически усредненным временем обработки произвольных данных. Поэтому максимально возможное время прохождения данных по конвейеру (латентность конвейера) в случае, когда все ступени СС-конвейера нужны для обработки текущей порции данных, может компенсироваться уменьшением латентности в остальных случаях.

Возможны следующие варианты уменьшения числа активно работающих ступеней СС-конвейера: обход одной или нескольких ступеней и мультиплексирование ступеней. Рассмотрим их подробнее.

### **3 Обход ступеней самосинхронного конвейера**

В качестве примера рассмотрим случай обхода двух ступеней конвейера, представленный на рис. 2. Реализацию обхода ступеней 2 и 3 обеспечивают  $n$ -разрядный элемент И (блок «&» на рис. 2), мультиплексор входов ступени 4 (MX), элемент ИЛИ-И и сигналы управления  $E_2$  и  $E_3$ . В ступень 1 добавляется логика детектирования условия, разрешающего обход ступеней 2 и 3, и формирования бифазного флага режима работы ( $U$  и  $UB$ ).

Рисунок 3 показывает реализацию триггера флага. Здесь  $L$  и  $LB$  — парафазный сигнал разрешения обхода ( $L = 1, LB = 0$ ) ступеней 2 и 3;  $E_1$  — сигнал разрешения переключения регистра ступени 1 в очередную фазу;  $E_2$  и  $E_3$  — сигналы управления ступеней 2 и 3;  $WE$  — индикатор триггера режима, одновременно служащий сигналом управления регистром ступени 1.

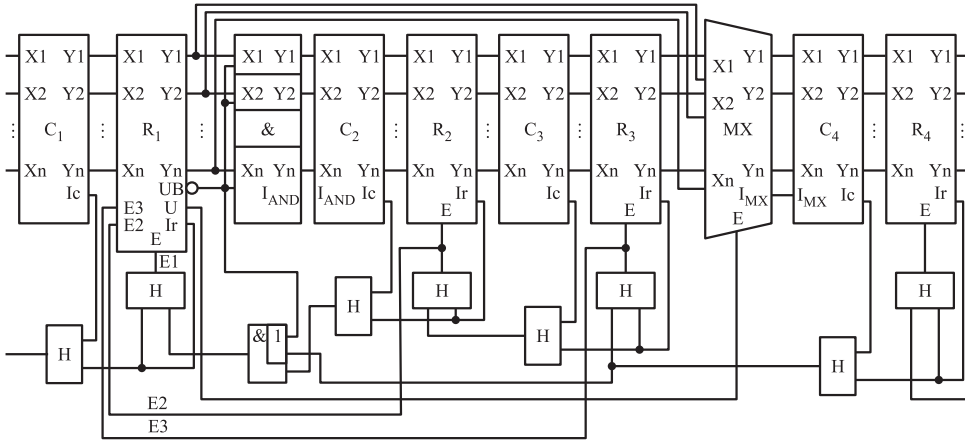


Рис. 2 Схема СС-конвейера с обходом ступеней 2 и 3

В общем случае маскируемые ступени  $(i + 1), \dots, (M + i + 1)$  постоянно находятся в спейсере, поскольку на их информационных входах поддерживается спейсерное состояние. В момент переключения в режим обхода завершение перехода  $i$ -й ступени в рабочую фазу задерживается на время переключения следующих  $(M + 1)$  ступеней в спейсер.

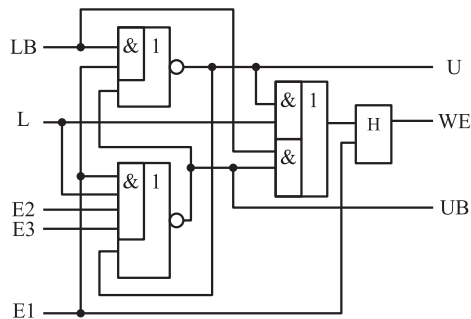


Рис. 3 Схема триггера режима и управления регистром ступени 1

#### 4 Мультиплексирование ветвей самосинхронного конвейера

В более общем случае последовательные ступеней не обходятся, а заменяются другими — параллельными  $L$  ступенями. Рассмотрим для примера конвейер на рис. 4. В нем ступень 2А включается в тракт обработки данных вместо ступеней 2 и 3 при выполнении условия  $U = 1$ . В сравнении со схемой на рис. 2 для организации корректной работы такого СС-конвейера требуется учет еще одного сигнала —  $E2A$ , входа управления ступенью 2А. Соответственно преобразуется и схема триггера режима (рис. 5).

Переключение с одной ветви конвейера на параллельную разрешается только после подтверждения перехода всех ступеней активной ветви в спейсер, что

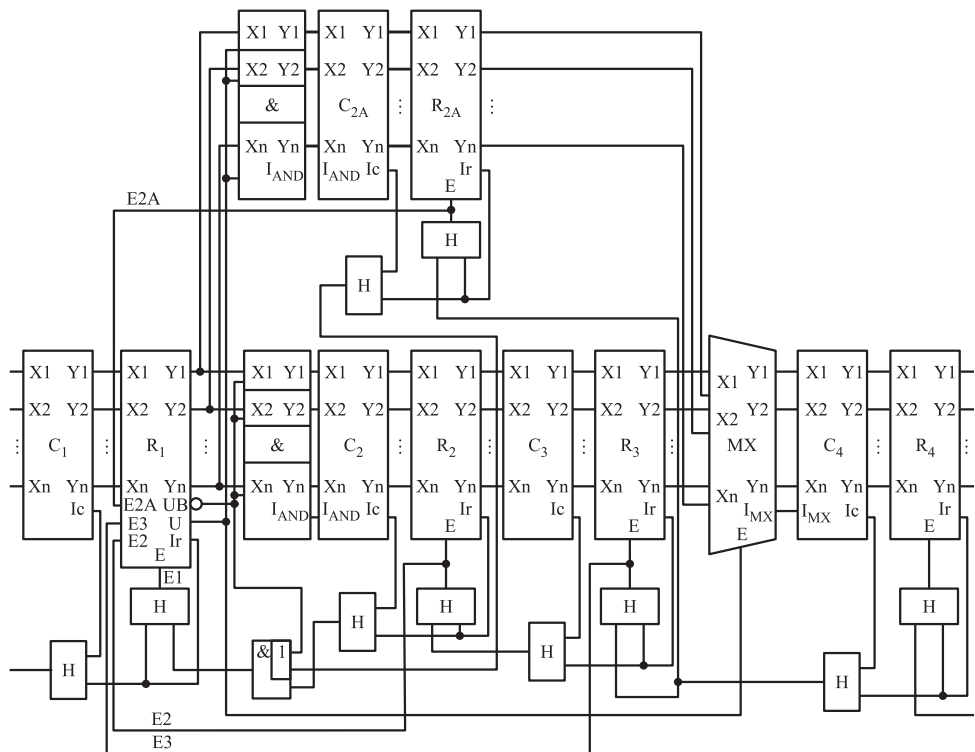


Рис. 4 Схема СС-конвейера с включением ступени 2А вместо ступеней 2 и 3

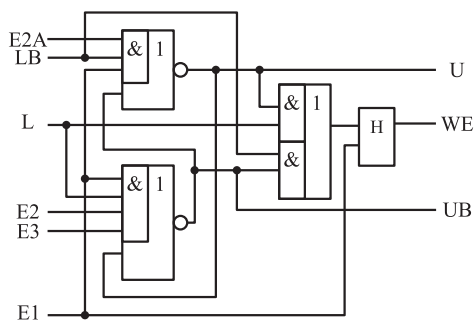


Рис. 5 Схема триггера режима и управления регистром ступени 1 для замены ступеней 2 и 3 ступенью 2А



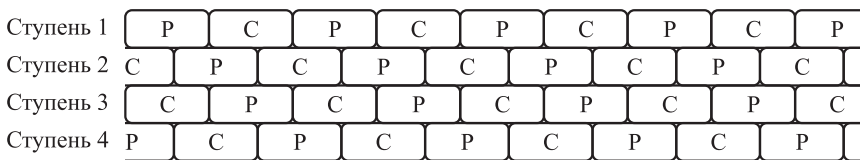
приводит к дополнительной задержке в работе конвейера и влияет на его латентность.

## 5 Латентность конвейера

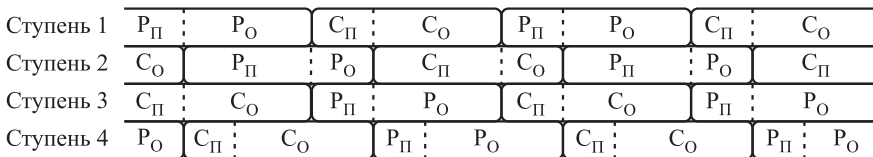
Целесообразность распараллеливания конвейера зависит от соотношения потерь в аппаратных затратах и выигрыша в производительности. Аппаратные затраты увеличиваются из-за появления параллельной ветви, обеспечивающей более быстрый способ обработки данных и усложнения схемы управления.

Производительность конвейера оценивается по латентности прохождения данных через него ( $T_{\Pi}$ ) и темпу (частоте) загрузки входных данных ( $F_3$ ), зависящему от длительности цикла работы (рабочая плюс спейсерная фазы)  $T_{\Pi}$  самой медленной ступени. Рисунок 6 иллюстрирует фазовую диаграмму работы типового сбалансированного конвейера, изображенного на рис. 1. Здесь «Р» и «С» — периоды пребывания ступени конвейера в рабочей и спейсерной фазе соответственно. Для простоты будем считать, что длительности обеих фаз всех ступеней одинаковы и равны  $T_{\Phi}$ . Тогда  $T_{\Pi} = 2T_{\Phi}$  и  $T_{\Pi} = 2,5T_{\Phi}$ . В общем случае для  $N$  ступеней конвейера  $T_{\Pi} \approx 0,25(N + 1)T_{\Pi}$ . Темп загрузки порций входных данных равен  $F_3 = 1/T_{\Pi}$ .

Производительность как синхронного, так и самосинхронного конвейера определяется быстродействием его самой сложной ступени. Пусть, например, ступень 2 конвейера самая сложная и медленная. Тогда диаграмма работы конвейера будет выглядеть так, как показано на рис. 7. Здесь каждая фаза условно разделена на две части: переключение в текущую фазу ( $P_{\Pi}$ ,  $C_{\Pi}$ ) и ожидание разрешения на переключение в противоположную фазу ( $P_O$ ,  $C_O$ ). Первая часть определяется сложностью данной ступени, а вторая — длительностью  $P_{\Pi}$  и  $C_{\Pi}$



**Рис. 6** Фазовая диаграмма сбалансированного СС-конвейера



**Рис. 7** Фазовая диаграмма несбалансированного СС-конвейера

самой сложной из остальных ступеней. Латентность такого конвейера описывается формулой  $T_{Л} \approx 0,25(N + 1)T_{Ц2}$ , а темп загрузки порций входных данных равен  $F_3 = 1/T_{Ц2}$ .

Замена  $M$  ступеней конвейера параллельными  $L$  ступенями (при простом обходе  $L = 0$ ) приводит к сокращению латентности конвейера, если разность суммарных латентностей  $M$  и  $L$  ступеней ( $T_{Л,M} - T_{Л,L}$ ) больше задержки, вносимой дополнительным ожиданием переключения  $M$  ступеней в спейсер при переключении в режим замены (обхода), блоками поддержания спейсера на входах первых ступеней после разветвления, мультиплексором перед первой после параллельных путей ступени конвейера и дополнительными элементами в схеме управления запросно-ответным взаимодействием ступеней: элементом ИЛИ-И и триггером режима, показанным на рис. 5. В первом приближении производительность конвейера на рис. 4 будет больше, чем у конвейера на рис. 1, если выполняется неравенство

$$T_{Л,2} + T_{Л,3} + 4T_{Г2} > T_{AND} + T_{MX} + T_{ТР} + T_{O21A} + P_M (T_{Л,2A} + T_{Г2}) + (1 - P_M) (T_{Л,2} + 1,5T_{Л,3} + 4T_{Г2} + 0,5T_{Л,4}), \quad (1)$$

где  $T_{Г2}$  — задержка двухвходового Г-триггера;  $T_{AND}$  — задержка блока «&»;  $T_{MX}$  — задержка блока МХ;  $T_{ТР}$  — задержка триггера режима;  $T_{O21A}$  — задержка логического элемента O21A;  $T_{Л,*}$  — латентности ступеней конвейера;  $P_M$  — вероятность появления условия замены ступеней 2 и 3 ступенью 2A. Пусть все ступени конвейера имеют примерно одинаковую латентность:

$$T_{Л,2} = T_{Л,2A} = T_{Л,3} = T_{Л,4} = T_{Л0}.$$

Тогда формула (1) перепишется в виде:

$$P_M (2T_{Л0} + 3T_{Г2}) > T_{Л0} + T_{AND} + T_{MX} + T_{ТР} + T_{O21A}.$$

При типовых задержках:  $T_{AND} = 20$  пс,  $T_{MX} = 30$  пс,  $T_{ТР} = 70$  пс,  $T_{Г2} = 40$  пс,  $T_{O21A} = 30$  пс,  $T_{Л0} = 300$  пс — для 65-нанометровой КМОП-технологии реализация режима замены ступеней 2 и 3 ступенью 2A окажется выгодной с точки зрения латентности конвейера при  $P_M > 0,63$ .

В случае простого обхода  $M$  ступеней рост производительности конвейера гарантируется при выполнении неравенства

$$MP_M (1,5T_{Л0} + 2T_{Г2}) > T_{AND} + T_{MX} + T_{ТР} + T_{O21A} + 0,5MT_{Л0}.$$

Тогда реализация режима обхода  $M$  ступеней приведет к повышению производительности конвейера при  $P_M > 0,28(1 + 1/M)$ . Обход одной ступени конвейера целесообразен при вероятности такого события более 0,56, а обход двух ступеней — уже при вероятности 0,43.

Таким образом, замедление конвейера в наихудшем случае из-за добавления в него аппаратуры обеспечения варьирования числа ступеней компенсируется его ускорением в остальных случаях, вероятности появления которых превышают некоторую величину.

Описанные схемотехнические реализации конвейера с маскируемыми и мультиплексируемыми ступенями были проверены на самосинхронность с помощью программы анализа цифровых схем на полумодулярность АСПЕКТ [5]. Проведенный анализ подтвердил корректность предлагаемых решений.

## 6 Заключение

Использование конвейерной организации СС-схемы позволяет повысить ее производительность аналогично синхронным схемам. Однако из-за двухфазного режима работы темп загрузки данных в СС-конвейер снижается.

Производительность СС-конвейера ограничивается быстродействием самых сложных ступеней. Время ожидания разрешения переключения в противоположную фазу работы остальных ступеней растет пропорционально времени переключения самой медленной ступени.

Построение СС-конвейера с опциональным мультиплексированием тракта обработки данных позволяет повысить среднюю производительность конвейера. Латентность СС-конвейера снижается, если условие разрешения замены двух ступеней одной параллельной ступенью выполняется в 63% циклов работы конвейера. Обход двух и более ступеней конвейера целесообразен при вероятности такого события более 0,43. Эффективность предложенного метода повышается с увеличением вероятности срабатывания условия обхода и числа ступеней на маскируемом участке конвейера.

## Литература

1. *Hennessy J. L., Patterson D. A.* Computer architecture: A quantitative approach. — 6th ed. — Morgan Kaufmann, 2019. 936 p.
2. *Varshavsky V. I., Kishinevsky M. A., Marakhovskiy V. B., Peschansky V. A.* Self-timed control of concurrent processes. — Kluwer Academic Pubs., 1990. 245 p. doi: 10.1007/978-94-009-0483-3.
3. *Stepchenkov Y. A., Kamenskih A. N., Diachenko Y. G., Rogdestvenski Y. V., Diachenko D. Y.* Improvement of the natural self-timed circuit tolerance to short-term soft errors // *Advances Science Technology Engineering Systems J.*, 2020. Vol. 5. No. 2. P. 44–56. doi: 10.25046/aj050206.
4. 754-2008 — IEEE Standard for Floating-Point Arithmetic. — IEEE Computer Society, 2008. doi: 10.1109/IEEE STD.2008.4610935.
5. *Рождественский Ю. В., Морозов Н. В., Рождественские А. В.* Подсистема событийного анализа самосинхронных схем АСПЕКТ // Проблемы разработки перспективных микро- и нанoeлектронных систем. — М.: ИППМ РАН, 2010. С. 26–31.

Поступила в редакцию 30.11.22

## SELF-TIMED PIPELINE WITH VARIABLE STAGE NUMBER

I. A. Sokolov, Yu. A. Stepchenkov, Yu. G. Diachenko, N. V. Morozov,  
and D. Yu. Diachenko

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

**Abstract:** The article considers the self-timed circuit’s performance improvement problem. As in synchronous circuits, an effective way to improve performance is to use a pipeline to implement multistage input data processing. The article analyzes possible options for dynamical reduction of the number of actively operating stages under certain conditions determined by the processed data value or an external signal. The estimates show that the efficiency of using an optionally variable number of pipeline stages depends on the number of bypassed stages and the probability of an event allowing this bypassing. In particular, replacing two successive pipeline stages with one parallel stage becomes expedient if it occurs in at least 63% of data processing operations and bypassing two or more stages reduces the average pipeline’s latency if it occurs in at least 43% of operations.

**Keywords:** self-timed circuit; pipeline; bypassing; multiplexing; latency; performance

**DOI:** 10.14357/08696527230101

### Acknowledgments

The research was supported by the Russian Science Foundation (project No. 22-19-00237).

### References

1. Hennessy, J. L., and D. A. Patterson. 2019. *Computer architecture: A quantitative approach*. 6th ed. Morgan Kaufmann. 936 p.
2. Varshavsky, V. I., M. A. Kishinevsky, V. B. Marakhovsky, and V. A. Peschansky. 1990. *Self-timed control of concurrent processes*. Kluwer Academic Publs. 245 p. doi: 10.1007/978-94-009-0483-3.
3. Stepchenkov, Y. A., A. N. Kamenskih, Y. G. Diachenko, Y. V. Rogdestvenski, and D. Y. Diachenko. 2020. Improvement of the natural self-timed circuit tolerance to short-term soft errors. *Advances Science Technology Engineering Systems J.* 5(2):44–56. doi: 10.25046/aj050206.
4. IEEE Computer Society. 2008. IEEE Standard for Floating-Point Arithmetic 754-2008. doi: 10.1109/IEEE STD.2008.4610935.
5. Rogdestvenski, Y. V., N. V. Morozov, and A. V. Rozhdestvenskene. 2010. Podsystema sobytiynogo analiza samosinkhronnyh skhem ASPEKT [ASPECT: A suite of self-timed event-driven analysis]. *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh*

*sistem* [Problems of the perspective micro- and nanoelectronic systems development].  
Moscow: IPPM RAN. 26–31.

*Received November 30, 2022*

## Contributors

**Sokolov Igor A.** (b. 1954) — Doctor of Science in technology, Academician of RAS, director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; isokolov@ipiran.ru

**Stepchenkov Yuri A.** (b. 1951) — Candidate of Science (PhD) in technology, head of department, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; YStepchenkov@ipiran.ru

**Diachenko Yuri G.** (b. 1958) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; diaura@mail.ru

**Morozov Nikolay V.** (b. 1956) — senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; NMorozov@ipiran.ru

**Diachenko Denis Y.** (b. 1987) — engineer-researcher, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; diaden87@gmail.com

## КОРПУСНЫЕ ДАННЫЕ ПРИ КОНТРАСТИВНОМ ИЗУЧЕНИИ ПУНКТУАЦИИ\*

*В. А. Нуриев<sup>1</sup>, М. Г. Кружков<sup>2</sup>*

**Аннотация:** Рассматриваются возможности использования корпусных данных при контрастивном исследовании пунктуации. Пунктуационные системы разных языков не всегда совпадают. В таких случаях пунктуационное оформление исходного (ИТ) и переводного текстов (ПТ) не может быть одинаковым. Проблемы межъязыковой пунктуационной асимметрии изучает контрастивная пунктуация — малоразработанная область лингвистики. Переводчику необходимо иметь четкое представление об этих межъязыковых расхождениях, чтобы оценить, насколько индивидуализирована пунктуационная манера в ИТ, и понять, как ее представить в ПТ. Для этого он может использовать различные информационные ресурсы, которые создаются в рамках информатики, компьютерной лингвистики и корпусного переводоведения. Один из таких ресурсов — надкорпусные базы данных (НБД), разрабатываемые в Федеральном исследовательском центре «Информатика и управление» Российской академии наук, позволяющие обрабатывать широкий массив параллельных текстов на разных языках и собирать эмпирические данные. Показано, как этот информационный ресурс можно использовать для контрастивного изучения пунктуации. В качестве иллюстративного материала привлекаются данные по нескольким пунктуационным знакам в русском и французском языке.

**Ключевые слова:** пунктуация; контрастивные исследования; перевод; художественный перевод; межъязыковая асимметрия; корпусное переводоведение; надкорпусная база данных; параллельный корпус; французский язык; русский язык

**DOI:** 10.14357/08696527230102

### 1 Введение

Знаки препинания образуют «буквенно-цифровой набор символов, которые используются для представления информации о структурных отношениях между составными элементами текста и включают в себя (во всяком случае, в европейских языках) запятую, точку с запятой, двоеточие, точку, скобки, кавычки

---

\*Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва) за счет гранта Российского научного фонда (проект 23-28-00548).

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, nurieff.v@gmail.com

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, magnit75@yandex.ru

и т. д.» [1, с. 17]. Их первостепенная функция — «привнести в текст большую структурную ясность и сигнализировать о семантических нюансах, которые иначе останутся скрытыми для читателя или, по меньшей мере, труднодоступными для его понимания» [2, с. 1].

Пунктуационные системы в разных языках могут не совпадать, что влечет за собой изменение пунктуационного оформления при переводе. Четкое представление о межъязыковой пунктуационной асимметрии помогает оценить, насколько индивидуализированы знаки препинания в ИТ, и принять решение о пунктуировании в переводе. Проблемы расхождений пунктуационных систем в разных языках изучает контрастивная пунктуация — малоразработанная лингвистическая область [3, с. 75–84]. Между тем о несовпадении пунктуационных тенденций от языка к языку писал еще Л. В. Щерба в первой половине XX в. [4]. Однако до настоящего времени контрастивные исследования в области пунктуации остаются спорадическими (см., например, [5–10]), а их методология страдает от недостаточной систематизации.

С одной стороны, пунктуация (в том числе в плане межъязыкового сопоставления) оставалась на периферии исследовательского интереса из-за своей креализованной семиотической природы. Основную функциональную нагрузку несет на себе графический сигнальный компонент, и поэтому знак препинания занимает промежуточное положение между единицами естественного языка и феноменами языка изобразительного. Отсюда сравнения пунктуации с нотной грамотой [11, 12]. С другой стороны, изучение пунктуации, как и любое исследование в области синтаксиса, предполагает высокие временные и трудовые издержки, так как сопряжено с обработкой больших текстовых массивов и комплексностью исследуемого объекта.

Дигитализация современной коммуникации привела к тому, что проблемы пунктуации переместились в сферу активных научных изысканий. Об этом, в частности, свидетельствуют монография 2015 г. семиотика Дж. Шайбла «Цифровой сдвиг: пунктуация и логика культуры» [13] и ряд других работ, затрагивающих, наравне с прочими аспектами, роль пунктуации при переводе [14–17]. Представляется, что для контрастивного исследования пунктуации наиболее подходят современные информационные корпусные инструменты, позволяющие автоматизированным образом обрабатывать большие массивы текстовых данных. В мировой практике для изучения пунктуации корпусный метод привлекается ограниченно. Во многом это обусловлено отсутствием общественно доступных электронных корпусов, содержащих сбалансированный и представительный массив переводных текстов. Сложности формирования таких информационных ресурсов и их предоставления в бесплатное общественное пользование сопряжены с проблемами правового характера (защита интеллектуальной собственности, авторского права и т. д.).

В качестве информационного ресурса для контрастивного исследования пунктуации может быть использована НБД, функционирующая на основе параллельных корпусов Национального корпуса русского языка (НКРЯ). Эти базы данных

разрабатываются в ФИЦ ИУ РАН. В них предусмотрены поисковые запросы для обработки пунктуационного наполнения текстов (см. разд. 2). Основной целью статьи, таким образом, ставится демонстрация возможностей информационных корпусных инструментов и полученных с их помощью данных при контрастивном исследовании пунктуации (см. разд. 3).

## **2 Надкорпусные базы данных и возможности обработки пунктуационного компонента**

Надкорпусные базы данных создаются для исследований в области контрастивной лингвистики, позволяя последовательно систематизировать и аннотировать эмпирические данные, заключенные в параллельных корпусах [18]. С помощью НБД можно анализировать большое число двуязычных примеров (извлеченных из ИТ и ПТ), сохраняя результаты этого анализа в унифицированном формате в виде аннотированных переводных соответствий. Такой подход дает возможность целенаправленно собирать факты речевой реализации различных языковых явлений и систематизировать их, разрабатывая при этом собственные специализированные классификации этих явлений. В результате создается обширный массив аннотированных переводных соответствий, оснащенный широкими поисковыми функциями.

Интерфейс НБД предоставляет доступ к загруженным в систему параллельным корпусам НКРЯ, на основе которых затем формируются аннотации переводных соответствий. Надкорпусные базы данных включают в себя поисковые функции, обеспечивающие первичный поиск текстовых фрагментов, где фиксируется то или иное языковое явление. Пользователи могут осуществлять поиск по словоформам и леммам слов с использованием подстановочных символов, в том числе с указанием различных диапазонов расстояний между искомыми элементами. Предусмотрена возможность одновременного задания поисковых запросов для фрагментов оригинала и перевода. Кроме того, при поиске можно указывать различные сочетания знаков препинания, присутствующие в искомых фрагментах. Последняя функция оказывается особенно полезной в тех случаях, когда исследуемые языковые явления в ИТ провоцируют употребление того или иного пунктуационного знака в ПТ. Например, при переводе коннекторов, выступающих показателями логико-семантических отношений, в качестве показателей соответствующих отношений нередко используются такие знаки пунктуации, как тире или двоеточие. В поисковой системе НБД могут использоваться любые сочетания знаков препинания, причем для некоторых из них, таких как тире и многоточие, предусмотрена возможность различного символического кодирования<sup>1</sup>.

---

<sup>1</sup>В настоящее время поиск по многоточию недоступен в поисковой системе НКРЯ; в НБД такой поиск возможен.



(Всего записей: 50)    перейти на страницу:  >

1 2 3 4 5

Номер пары	Оригинальный текст	Перевод	Построено
269	Pour l'éloge un : magicien.	В похвалу — одно: волшебник.	<a href="#">Выбрать</a>
443	Déserteur pour les uns, suspect aux autres : c'est la solitude de Calchas.	Для одних — предатель, для других — подозрительный тип: одиночество Калханта.	<a href="#">Выбрать</a>
503	Au dixième rappel Hofmannsthal se pencha vers Diaghilev : « J'aurais préféré avoir un scandale », lui dit-il.	Когда актеров вызвали в десятый раз, Гофмансталь наклонился к Дягилеву. «Я бы предпочел скандал», — сказал он.	<a href="#">Выбрать</a>
563	Et si le monde insiste : « Mon arme secrète est une tradition d'anarchie. »	А если будут настаивать, она скажет: «Мое секретное оружие — анархическая традиция».	<a href="#">Выбрать</a>

Рис. 1 Фрагмент выдачи поискового запроса (тире в переводе)

(Всего записей: 12)    перейти на страницу:  >

1 2

Номер пары	Оригинальный текст	Перевод	Построено
1507	Bizarre vieillard qui ferme les yeux, hoche la tête, prend un air de suivre les paroles et se répète à soi-même : « Je suis le cancre du collège. »	Странного вида старик, который, закрыв глаза, качает головой и делает вид, что внимательно слушает, а сам думает: «Двоечник я, двоечник...»	<a href="#">Выбрать</a>

Рис. 2 Фрагмент выдачи поискового запроса (многоточие в переводе)

На рис. 1 представлены несколько найденных с помощью поисковой системы НБД примеров, где в ИТ — двоеточие, а в ПТ — двоеточие и тире.

На рис. 2 показан пример, где точка в ИТ при переводе преобразуется в многоточие.

Находя и аннотируя подобные примеры, лингвисты генерируют новые знания, выражающиеся как в новых системах классификации для исследуемых явлений, так и в результатах анализа обобщенных данных по различным межъязыковым соответствиям. Ниже представлены два примера<sup>1</sup>.

(1) «C'est être un velléitaire, un traître, un acrobate, un fantaisiste. Pour l'éloge un: magicien».

«А это значит: слабак, иуда, ловкач, сумасброд. В похвалу — одно: волшебник» (М. Аннинская).

«А это означает прослыть слабовольным, изменником, акробатом, сумасбродом. Или в виде похвалы — волшебником» (Л. Цывьян).

При переводе (1) в обоих вариантах поставлено тире, но первый переводчик сохраняет и двоеточие, создавая нетривиальную пунктуационную ситуацию — заставляя тире и двоеточие конкурировать. Это помогает навести фокус внимания

<sup>1</sup>Источник примеров (1) и (2) — «La difficulté d'être» (1947 г.) Ж. Кокто и два русскоязычных перевода: «Тяжесть бытия» и «Трудность бытия», сделанные в Л. М. Цывьяном и М. Л. Аннинской и опубликованные в 2003 и 2005 гг. соответственно.

на слово, находящееся между двумя знаками препинания. Второй переводчик строит синтаксическую структуру, приспособленную для постановки тире.

(2) «Bizarre vieillard qui ferme les yeux, hoche la tête, prend un air de suivre les paroles et se répète à soi-même : “Je suis le cancre du collège.”»

«Странного вида старик, который, закрыв глаза, качает головой и делает вид, что внимательно слушает, а сам думает: “Двоечник я, двоечник...”» (М. Аннинская).

«Глупый старик, который прикрывает глаза, кивает головой, всем видом изображает, будто ловит каждое слово, а на самом деле мысленно твердит, адресуясь к себе: “Тупица, двоечник”» (Л. Цывьян).

Замена конечной точки многоточием в (2) у первого переводчика указывает на то, что этот пунктуационный знак может иметь особенную функциональную нагрузку — выступать средством углубления эмоциональной перспективы высказывания. В данном случае подразумевается наличие эмоционального подтекста, связанного с самокритикой, и многоточие может сигнализировать о сожалении речевого субъекта по поводу сложившегося положения дел.

### 3 Экспериментальные данные

Объект этого исследования — два пунктуационных знака в русском и французском языке (двоеточие и многоточие). Таблицы 1 и 2 содержат данные, в числовом эквиваленте отражающие их употребление при переводе с французского на русский и обратно. В табл. 1 представлены данные по пунктуированию в четырех французских произведениях и их переводах на русский (для «La difficulté d'être» Ж. Кокто имеются два перевода). В табл. 2 приводятся данные по пунктуированию в двух русскоязычных произведениях и их переводах на французский (для «Шинели» Н. В. Гоголя имеются четыре перевода).

**Таблица 1** Пунктуирование в ИТ и его переводе (направление перевода: французский → русский)

Знак	<i>La difficulté d'être</i> , 1947 / «Тяжесть бытия», 2003; «Трудность бытия», 2005		
	Ж. Кокто	Л. М. Цывьян	М. Л. Аннинская
:	88	102	151
...	7	6	6

Знак	<i>99 francs</i> , 1997 / «99 франков», 2000		<i>Petit prince</i> , 1942 / «Маленький принц», 1959		<i>Quartier perdu</i> , 1985 / «Утраченный мир», 1989	
	Ф. Бергедер	И. Я. Волевич	А. де Сент-Экзюпери	Н. Галь	П. Модино	Ю. Я. Яхнина
:	542	576	157	141	148	127
...	114	143	189	165	410	395

**Таблица 2** Пунктуирование в ИТ и его переводе (направление перевода: русский → французский)

Знак	«Шинель», 1842 / <i>Le Manteau</i> , 1925, 1938, 1998, 2007					«Теллурия», 2013 / <i>Telluria</i> , 2017	
	Н. В. Гоголь	Б. де Шлёцер	А. Монго	Б. Крез	А. Марковиц	В. Г. Сорокин	А. Кольдефи-Фокар
:	76	85	74	95	66	377	497
...	19	22	27	21	21	500	500

Анализ данных, представленных в табл. 1 и 2, не позволяет выявить четких тенденций в употреблении двоеточия и многоточия при переводе с французского языка на русский и обратно. Объем корпуса задействованных параллельных текстов для этого пока недостаточен. Вместе с тем эти данные позволяют сделать несколько предварительных выводов. Например, исходное пунктуационное действие и пунктуационная установка, его мотивирующая, очевидным образом осмысляются переводчиком, поскольку число знаков препинания в оригинале и переводе совпадает не всегда. Совпадают отдельные количественные характеристики. Так, А. Кольдефи-Фокар в переводе «Теллурии» В. Г. Сорокина употребляет столько же многоточий, сколько и в ИТ (500 шт.), что указывает на установку, запрещающую отклонение от заданного ИТ пунктуационного контура, но этот запрет не распространяется на двоеточие. В четырех переводах «Шинели» Н. В. Гоголя нигде не сохраняется исходное число двоеточий и многоточий. Эти четыре перевода отражают совершенно разные пунктуационные предпочтения переводчиков. То же самое можно сказать и о переводах в направлении французский → русский, где в отношении обоих знаков переводчики используют пунктуационные возможности родного языка неодинаково. Примечательно, что большая часть обработанных переводов (за исключением «99 франков» И. Я. Волевич) характеризуется «сдержанной» постановкой многоточия: число многоточий либо совпадает в ИТ и ПТ, либо незначительно увеличивается/уменьшается в ПТ по сравнению с ИТ.

## 4 Заключение

В статье показано, как НБД — специализированный информационный ресурс, разрабатываемый на стыке информатики, компьютерной лингвистики и корпусного переводоведения, — можно использовать для контрастивного изучения пунктуации. Эмпирическим материалом послужили данные по употреблению двоеточия и многоточия при переводе. Полученные выводы носят предварительный характер, указывая на необходимость дальнейших исследований в этой области с применением корпусных инструментов, что позволит систематизировать методологию изучения и выявить более четкие тенденции

в функционировании знаков препинания в контактирующих при переводе языках.

## Литература

1. *Nunberg G.* The linguistics of punctuation. — Stanford, CA, USA: Center for the Study of Language and Information, Leland Stanford Junior University, 1990. 160 p.
2. *Parkes M. B.* Pause and effect: An introduction to the history of punctuation in the West. — Aldershot: Scholar Press, 1992. 343 p.
3. *Сигал К. Я.* К основаниям контрастивной пунктуации // Проблемы теории синтаксиса. — М.: Ключ-С, 2012. С. 75–84.
4. *Щерба Л.* Пунктуация // Литературная энциклопедия. — М.: ОГИЗ РСФСР: Гос. ин-т. «Сов. энцикл.», 1935. Т. 9. Стб. 366–370.
5. *Bystrova-McIntyre T.* Looking at the overlooked: A corpora study of punctuation use in Russian and English // Transl. Interpret. Stu., 2007. Vol. 2. Iss. 1. P. 137–162.
6. *Сигал К. Я.* Пунктуация как средство создания эмоционального подтекста (на материале рассказа М. А. Шолохова «Судьба человека» и его переводов на английский язык) // Известия РАН. Серия литературы и языка, 2014. Т. 73. № 6. С. 38–50.
7. *Wang C.* Decoding and encoding the discourse meaning of punctuation: A perspective from English-to-Chinese translation // Babel — Amsterdam, 2018. Vol. 64. Iss. 2. P. 225–249. doi: 10.1075/babel.00032.wan.
8. *Wollin L.* Punctuation: Providing the setting for translation? // Stud. Neophilol., 2018. Vol. 90. No. S1. P. 37–49.
9. *Fuhrhop N., Hettwer C.* Comparing word marks // Vergleichende interpunktion — comparative punctuation / Eds. P. Rössler, P. Besl, A. Saller. — Berlin, Boston: De Gruyter, 2021. P. 263–288. doi: 10.1515/9783110756319-012.
10. *Pecorari F., Longo F.* The ellipsis and the dash in Italian and English: A contrastive perspective // Vergleichende interpunktion — comparative punctuation / Eds. P. Rössler, P. Besl, A. Saller. — Berlin, Boston: De Gruyter, 2021. P. 289–314. doi: 10.1515/9783110756319-013.
11. *Adorno Th. W.* Satzzeichen // Noten zur Literatur / Ed. R. Tiedemann. — Frankfurt/M.: Suhrkamp, 1989. P. 106–113.
12. *Карпов В. И.* Информационная эстетика Макса Бензе и теория текста: к истокам одной терминологической традиции в немецком языкознании // Критика и семиотика, 2022. № 1. С. 143–157. doi: 10.25205/2307-1737-2022-1-143-157.
13. *Scheible J.* Digital shift: The cultural logic of punctuation. — Minneapolis, MN, USA: University of Minnesota Press, 2015. 176 p.
14. *Lukeman N.* A dash of style: The art and mastery of punctuation. — New York, NY, USA: W. W. Norton & Co., 2006. 206 p.
15. *Mogahed M.* Punctuation marks make a difference in translation: Practical examples, 2012. <https://files.eric.ed.gov/fulltext/ED533736.pdf>.
16. *Brusasco P., Corino E.* Translating punctuation // Translating and comparing languages: Corpus-based insights / Eds. S. Granger, M.-A. Lefer. — Louvain-la-Neuve, Belgique: Presses Universitaires de Louvain, 2020. P. 101–122.
17. *Youdale R.* Using computers in the translation of literary style: Challenges and opportunities. — London, New York: Routledge, 2020. 242 p.

18. Зализняк Анна А., Зацман И. М., Инькова О. Ю., Кружков М. Г. Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика: Тр. VII Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.

Поступила в редакцию 31.01.23

---

---

## THE PARALLEL CORPORA PERSPECTIVE ON STUDYING CONTRASTIVE PUNCTUATION

*V. A. Nuriev and M. G. Kruzchkov*

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

**Abstract:** The paper presents the parallel corpora perspective on studying contrastive punctuation. Different languages often use the same punctuation signs in a quite different manner which challenges a translator. Differences in punctuation systems of various languages are studied in the field of contrastive punctuation, a rather underdeveloped branch of linguistics. The translator has to have an accurate understanding of these differences in order to see how particular and individualized the punctuation is in the source text so as to decide how to render it in the target language. To this end, the translator can take advantage of various information resources that result from the joint efforts of information science, computer linguistics, and corpus-based translation studies. One of such resources is supracorpora databases developed at the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences. They store parallel texts (in different languages) from the Russian National Corpus and allow users to collect massive empirical data. The paper shows how this information resource can be used to research contrastive punctuation. Also, it draws some preliminary conclusions on functioning of several punctuation signs in French and Russian.

**Keywords:** punctuation; contrastive studies; translation; literary translation; asymmetry between languages; corpus-based translation studies; supracorpora database; parallel corpus; French; Russian

**DOI:** 10.14357/08696527230102

### Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Информатика”) of FRC CSC RAS (Moscow). The research was supported by the Russian Science Foundation (project No. 23-28-00548).

### References

1. Nunberg, G. 1990. *The linguistics of punctuation*. Stanford, CA: Center for the Study of Language and Information, Leland Stanford Junior University. 160 p.

2. Parkes, M. B. 1992. *Pause and effect: An introduction to the history of punctuation in the west*. Aldershot: Scholar Press. 343 p.
3. Seagal, K. Ya. 2012. K osnovaniyam kontrastivnoy punktuatsii [On the foundations of contrastive punctuation]. *Problemy teorii sintaksisa* [Problems of the theory of syntax]. Moscow: Klyuch-S. 75–84.
4. Shcherba, L. 1935. Punktuatsiya [Punctuation]. *Literaturnaya entsiklopediya* [Literary encyclopedia]. Moscow: OGIZ RSFSR. 9:366–370.
5. Bystrova-McIntyre, T. 2007. Looking at the overlooked: A corpora study of punctuation use in Russian and English. *Transl. Interpret. Stu.* 2(1):137–162.
6. Seagal, K. Ya. 2014. Punktuatsiya kak sredstvo sozdaniya emotsional'nogo podteksta (na materiale rasskaza M. A. Sholokhova "Sud'ba cheloveka" i ego perevodov na angliyskiy yazyk) [Punctuation as a means of creating of emotional subtext (on the material of the M. A. Sholokhov's story "The fate of a man" and its translations into English)]. *Bulletin of the Russian Academy of Sciences: Studies in Literature and Language* 73(6):38–50.
7. Wang, C. 2018. Decoding and encoding the discourse meaning of punctuation: A perspective from English-to-Chinese translation. *Babel — Amsterdam* 64(2):225–249. doi: 10.1075/babel.00032.wan.
8. Wollin, L. 2018. Punctuation: Providing the setting for translation? *Stud. Neophilol.* 90(S1):37–49.
9. Fuhrhop, N., and C. Hettwer. 2021. Comparing word marks. *Vergleichende Interpunktion — comparative punctuation*. Eds. P. Rössler, P. Besl, and A. Saller. Berlin, Boston: De Gruyter. 263–288. doi: 10.1515/9783110756319-012.
10. Pecorari, F., and F. Longo. 2021. The ellipsis and the dash in Italian and English: A contrastive perspective. *Vergleichende Interpunktion — comparative punctuation*. Eds. P. Rössler, P. Besl, and A. Saller. Berlin, Boston: De Gruyter. 289–314. doi: 10.1515/9783110756319-013.
11. Adorno, Th. W. 1989. Satzzeichen. *Noten zur Literatur*. Ed. R. Tiedemann. Frankfurt/M.: Suhrkamp. 106–113.
12. Karpov, V. I. 2022. Informatsionnaya estetika Maksa Benze i teoriya teksta: K istokam odnoy terminologicheskoy traditsii v nemetskom yazykoznanii [Max Bense's information aesthetics and text theory: To the origins of a terminological tradition in German linguistics]. *Kritika i semiotika* [Critique and Semiotics]. 1:143–157. doi: 10.25205/2307-1737-2022-1-143-157.
13. Scheible, J. 2015. *Digital shift: The cultural logic of punctuation*. Minneapolis, MN: University of Minnesota Press. 176 p.
14. Lukeman, N. 2006. *A dash of style: The art and mastery of punctuation*. New York, NY: W. W. Norton & Co. 206 p.
15. Mogahed, M. 2012. Punctuation marks make a difference in translation: Practical examples. Available at: <https://files.eric.ed.gov/fulltext/ED533736.pdf> (accessed March 13, 2023).
16. Brusasco, P., and E. Corino. 2020. Translating punctuation. *Translating and comparing languages: Corpus-based insights*. Eds. S. Granger and M.-A. Lefer. Louvain-la-Neuve, Belgique: Presses Universitaires de Louvain. 101–122.
17. Youdale, R. 2020. *Using computers in the translation of literary style: Challenges and opportunities*. London, New York: Routledge. 242 p.

18. Zaliznyak, Anna A., I. M. Zatsman, O. Yu. Inkova, and M. G. KruzHKov. 2015. Nadkorpornye bazy dannyykh kak lingvisticheskiy resurs [Supracorpora databases as linguistic resource]. *7th Conference (International) on Corpus Linguistics Proceedings*. St. Petersburg: St. Petersburg State University Pubs. 211–218.

*Received January 31, 2023*

## **Contributors**

**Nuriev Vitaly A.** (b. 1980) — Doctor of Science in philology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; nurieff.v@gmail.com

**KruzHKov Mikhail G.** (b. 1975) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

## ИНТЕГРАЦИОННЫЕ ВОЗМОЖНОСТИ НАДКОРПУСНЫХ БАЗ ДАННЫХ\*

*А. А. Дурново<sup>1</sup>, О. Ю. Инькова<sup>2</sup>, В. А. Нуриев<sup>3</sup>*

**Аннотация:** Описываются интеграционные возможности надкорпусных баз данных, разработанных в ФИЦ ИУ РАН. Продемонстрировано, как три базы данных — Надкорпусная база данных иерархии логико-семантических отношений (НБДИ ЛСО), База данных параллельных текстов (БДТ) и Надкорпусная база данных коннекторов (НБДК) — интегрированы между собой. Информационная система иерархии логико-семантических отношений (ИСИЛСО) использует специально спроектированную базу данных (НБДИ ЛСО), в которой аннотации логико-семантических отношений представлены в виде деревьев, т. е. ориентированных связных ациклических графов, где узлы содержат данные, а ребра отображают отношение подчиненности между узлами. Наряду с НБДИ ЛСО ИСИЛСО использует данные из БДТ и НБДК. Такая интеграция позволяет комбинировать методологические наработки информатики, контрастивной и корпусной лингвистики, теории и практики перевода, не упуская из виду факторы, которые могут отрицательно сказаться на валидности и надежности конечных данных.

**Ключевые слова:** надкорпусная база данных; интеграция баз данных; многоязычный корпус; параллельный корпус; корпусные информационные ресурсы; переводоведение; контрастивная лингвистика; машинный перевод

**DOI:** 10.14357/08696527230103

### 1 Введение

Контрастивная лингвистика, переводоведение и практика (машинного и осуществляемого человеком) перевода имеют очевидные точки пересечения: по меньшей мере каждая из указанных областей связана с обработкой двух и более естественных языков. Кроме того, сейчас все три области в своей методологии

---

\*Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва) в рамках швейцарско-российского проекта «Методология аннотирования в надкорпусной базе данных коннекторов» по гранту Швейцарского государственного секретариата по образованию, исследованиям и инновациям.

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, duralex49@mail.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Женевский университет, olyainkova@yandex.ru

<sup>3</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, nurieff.v@gmail.com



ориентируются на корпусные методы исследования и получения данных. В машинном переводе многоязычный корпус служит источником параллельных текстов, используемых в качестве тренировочных выборок для переводных моделей. Переводоведы используют корпусные данные для описания процесса перевода и его теоретического моделирования (с учетом переводческих стратегий и особенностей переводимого текста). Практикующий переводчик из параллельного корпуса получает референтные образцы, и в этом смысле подобные информационные ресурсы максимально сближаются с двуязычным словарем, значительно превосходя его по широте наполнения и удобству пользования. В контрастивной лингвистике корпусные данные служат эмпирическим базисом, позволяющим упрочить теоретические изыскания и более рельефно показать межъязыковые различия, не попадавшие до этого в фокус исследования (подробнее о взаимодействии трех областей см. [1, с. v–vi, 1]).

Несмотря на существующие точки соприкосновения, методологический взаимодействие между тремя областями не является паритетным. Развитие корпусной парадигмы его ускорило, не обеспечив при этом равноправного взаимовлияния вовлеченных дисциплин, теории и практики. Переводоведение активно пользуется наработками контрастивной лингвистики, в то время как последняя гораздо реже обращается к опыту, накопленному в теории перевода, проявляя интерес скорее к переводам как объекту анализа, но не к методологии корпусной работы с ними. Поэтому в контрастивных исследованиях могут не учитываться некоторые особенности переводного процесса, определяющие валидность и надежность получаемых данных (см., например, [2, 3]). Так, при использовании параллельного корпуса не всегда учитываются факторы, способные существенно исказить результирующие данные, — направление перевода, индивидуальная переводческая стратегия, особенности переводного текста (жанр, регистр, время возникновения) и т. д.

За последние несколько лет вышел ряд работ, где предпринимается попытка исправить сложившееся положение [4–9], в том числе публикации исследователей, осуществляющих свою научную деятельность в ФИЦ ИУ РАН [10, 11].

В данной статье рассматривается НБДИ ЛСО и возможности ее интеграции с другими базами данных, функционирующими на основе параллельных корпусов. Это один из информационных корпусных ресурсов, позволяющих успешно комбинировать наработки контрастивной лингвистики и теории перевода, не упуская из виду факторы, которые могут негативно повлиять на валидность и надежность конечных данных.

Надкорпусная база данных иерархии логико-семантических отношений<sup>1</sup> создана в ФИЦ ИУ РАН, чтобы расширить возможности аннотирования логико-семантических отношений и их показателей, имеющиеся в НБДК, также созданной в ФИЦ ИУ РАН [13]. Надкорпусная база данных иерархии логико-семантических отношений использует тексты параллельных подкорпусов

---

<sup>1</sup> Архитектура НБДИ ЛСО описана в работе [12].

Национального корпуса русского языка [14] общим объемом свыше 11 млн словоупотреблений, а также аннотации НБДК.

## 2 Интеграция баз данных НБДИ ЛСО, БДТ и НБДК

Информационная система иерархии логико-семантических отношений, обсуждаемая в настоящей работе, использует специально спроектированную базу данных (НБДИ ЛСО), в которой аннотации логико-семантических отношений представлены в виде деревьев, т. е. ориентированных связанных ациклических графов, где узлы (вершины) содержат данные, а ребра отображают отношение подчиненности между узлами.

Информационная система иерархии логико-семантических отношений помимо НБДИ ЛСО использует информацию из БДТ и НБДК.

База данных параллельных текстов служит источником параллельных текстов, на основе которых создаются аннотации. Основной единицей хранения текстов в БДТ служит *трансляция*. Трансляция представляет собой упорядоченную последовательность пар, где первым элементом выступает *фрагмент* (одно или несколько предложений) текста оригинала, а вторым — соответствующий ему фрагмент текста перевода:  $\langle\langle o_0, t_0 \rangle, \langle o_1, t_1 \rangle, \langle o_2, t_2 \rangle, \dots, \langle o_{n-1}, t_{n-1} \rangle\rangle$ . Здесь  $n$  — число пар фрагментов в трансляции,  $o_0, o_1$  и  $o_{n-1}$  — первый, второй и последний фрагменты текста оригинала;  $t_0, t_1$  и  $t_{n-1}$  — первый, второй и последний фрагменты текста перевода. Последовательность  $\langle o_0, o_1, o_2, \dots, o_{n-1} \rangle$  соответствует полному тексту оригинала, а последовательность  $\langle t_0, t_1, t_2, \dots, t_{n-1} \rangle$  — полному тексту перевода. Пары вида  $\langle o_i, t_i \rangle$  будем называть парами фрагментов, или просто *парами*.

Связь НБДИ ЛСО с НБДК более тесная прежде всего потому, что одним из важнейших типов данных в НБДИ ЛСО, как и в НБДК, выступают коннекторы. Кроме того, множество пользователей, имеющих возможность работать с НБДИ ЛСО, совпадает с множеством пользователей НБДК.

Главным типом объектов в НБДК служат *аннотации*. Важнейшей частью аннотации является пара (в математическом смысле) вида  $\langle g_0, g_t \rangle$ , где  $g_0$  — *контекст* оригинала;  $g_t$  — контекст перевода. Контекст, в свою очередь, упрощенно может быть представлен как тройка вида  $\langle w, t, p \rangle$ , где  $w$  — последовательность словоформ фрагмента (нескольких соседних фрагментов);  $t$  — тип речевой реализации коннектора, т. е. та форма, в которой коннектор зафиксирован в аннотируемом контексте<sup>1</sup>;  $p$  — набор свойств, характеризующих коннектор и контекст

---

<sup>1</sup> Речь идет о таких, например, случаях, как речевые реализации *да и еще* и *да притом еще* в следующих примерах: А я вот представлю... совершила я что-то эдакое... мне и так плохо, *да и еще* самые близкие и дорогие во мне разочарованы (родители) — к кому бежать, от кого ждать помощи? [Наши дети: Подростки (2004)]; — Из русского окопа ответили одним словом, *да притом еще* очень коротким [Василий Гроссман. Жизнь и судьба. Ч. 3 (1960)]. Ср. кластер речевых реализаций коннектора *да*, зафиксированных в НБДК: *да, да и, да еще и, да и еще, да притом еще, да притом еще и* и др. Подробнее о формальной вариативности коннекторов см. [15].

его употребления. При этом фрагменты обоих контекстов должны относиться к одной и той же паре (парам — в случае нескольких соседних фрагментов).

Узлы дерева в НБДИ ЛСО бывают двух типов: Т-узлы, содержащие текст, и С-узлы, содержащие коннектор. Корневым узлом дерева всегда является Т-узел. Дочерними узлами Т-узлов могут быть только С-узлы, причем у каждого Т-узла может быть не более одного дочернего, тогда как дочерними узлами С-узлов могут быть только Т-узлы, причем для каждого С-узла в количестве не менее двух.

В результате тесных связей НБДИ ЛСО с БДТ и НБДК из восьми таблиц, имеющих в НБДИ ЛСО, шесть имеют поля, которые тем или иным образом ссылаются на БДТ или НБДК. Так, в таблице Structure, которая содержит информацию о дереве в целом, имеются поля TranslationId (идентификатор трансляции в БДТ), ParaNumStart и ParaNumEnd (номера соответственно первой и последней пар из трансляции, определяющих диапазон текста, в пределах которого строится дерево), FromTransl (строится ли дерево на основе оригинала или, напротив, перевода) и LangId (идентификатор языка текста в БДТ). В таблице Context, которая соответствует Т-узлам дерева, имеются поля WNumTextStart и WNumTextEnd (порядковые номера первой и последней словоформ оригинального или переводного текста трансляции, которые определяют диапазон текста, являющегося контекстом данного Т-узла). В таблице Link, соответствующей С-узлам дерева, есть поле LGFTypeID (идентификатор типа речевой реализации коннектора в НБДК). Таблица Link\_LGFProp, которая задает множество свойств из кластера «Отношения», характеризующих коннектор данного С-узла, имеет поле LGFPropId (идентификатор свойства из НБДК). Таблица Portion, которая соответствует отрезкам текста в контексте Т-узла (если текст не сплошной, а имеет пропуски), содержит поля WNumTextStart и WNumTextEnd (выполняющие ту же функцию, что и одноименные поля таблицы Context). Наконец, таблица History, содержащая информацию о всех состояниях каждого дерева в разные периоды времени, имеет поле UserId (идентификатор пользователя из НБДК).

Далее будут приведены примеры совместного использования нескольких баз в рамках одной и той же операции с данными.

### 3 Примеры совместного использования нескольких баз

#### 3.1 Вход в систему

Доступ к ИСИЛСО имеют только пользователи, зарегистрированные в НБДК. Для входа в систему необходимо ввести те же логин и пароль, которые используются для подключения к пользовательскому интерфейсу для работы с НБДК.

Приведем пример запроса к НБДИ ЛСО на языке MS SQL, где проверяется корректность введенных пользователем логина и пароля:

```
select top 1 [Id] from [LingvoConnector].[dbo].[Users]
where [Login] = N'Ivanov' and [Password] = N'287@307'
```

Здесь и далее в примерах LingvoText — имя БДТ, а LingvoConnector — имя НБДК.

### 3.2 Создание дерева

При необходимости создания нового дерева у пользователя есть три возможности:

- (1) создать дерево по идентификатору аннотации из НБДК;
- (2) создать дерево по идентификатору пары из БДТ;
- (3) создать дерево по отрывку текста.

В каждом случае в НБДТ происходит поиск (в первом случае с использованием информации, хранящейся в НБДК) хотя бы одной пары, удовлетворяющей требуемому критерию. Если такая пара будет найдена, текст ее левого или правого (в зависимости от выбранного языка) фрагмента станет корневым контекстом нового дерева. При создании дерева корневой контекст может быть расширен за счет фрагментов соседних пар. Это бывает необходимо в том случае, когда корневой контекст недостаточен для определения свойств, характеризующих данное употребление коннектора [16].

В результате создается вырожденное дерево с единственным T-узлом в качестве корневого.

Приведем пример запроса к НБДИ ЛСО на языке MS SQL, где требуется найти в БДТ пару по идентификатору аннотации в НБДК, равному 19711:

```
with [Aux] as
(
select distinct
    [t].[Id] as [TranslationId],
    [a].[Name] as [AuthorName],
    [b].[Title] as [BookTitle],
    [p].[ParaNum] as [ParaNum],
    [fo].[Text] as [OrigText],
    [ft].[Text] as [TranslText]
from [LingvoConnector].[dbo].[Equivalence] [e]
join [LingvoConnector].[dbo].[LGF] [go] on [go].[Id] = [e].[LGFOrigId]
join [LingvoConnector].[dbo].[LGF_Word] [gwo] on [gwo].[LGFId] = [go].[Id]
join [LingvoText].[dbo].[WordInText] [wo] on [wo].[Id] = [gwo].[WordId]
join [LingvoText].[dbo].[Fragment] [fo] on [fo].[Id] = [wo].[FragmentId]
join [LingvoText].[dbo].[Para] [p] on [p].[FragOrigId] = [fo].[Id]
join [LingvoText].[dbo].[Fragment] [ft] on [ft].[Id] = [p].[FragTranslId]
join [LingvoText].[dbo].[Translation] [t] on [t].[Id] = [p].[TranslationId]
and [t].[LangId] in (2, 9)
join [LingvoText].[dbo].[Book] [b] on [b].[Id] = [t].[BookId]
and [b].[LangId] = 1
```

```

join [LingvoText].[dbo].[Author] [a] on [a].[Id] = [b].[AuthorId]
where [e].[Id] = 19711
)
select top 1
  [TranslationId],
  [AuthorName],
  [BookTitle],
  [ParaNum],
  [OrigText],
  [TranslText]
from [Aux]
where [TranslationId] = (select min([TranslationId]) from [Aux])

```

Ответ на запрос содержит идентификатор трансляции, имя автора, название произведения, порядковый номер пары в трансляции, а также тексты фрагментов.

### 3.3 Дальнейшее развитие дерева

В ходе дальнейшего построения дерева происходит разделение контекстов на составляющие с появлением узлов, подчиненных корневому. При этом возникают узлы, содержащие коннектор, т. е. С-узлы. Создание каждого С-узла вызывает необходимость поиска информации в НБДК, где содержатся данные о речевых реализациях коннекторов, а также об их свойствах из кластера «Отношения».

Приведем пример запроса к НБДИ ЛСО на языке MS SQL, где требуется получить в НБДК список типов коннекторов для русского языка:

```

select [Id], [Code] from [LingvoConnector].[dbo].[LGFTType]
where [LangId] = 1 order by [Id]

```

Более того, при отсутствии в НБДК необходимой речевой реализации коннектора пользователь имеет возможность ее создать в НБДК. То же касается и свойств коннекторов из кластера «Отношения». Из этого следует, что ИСИЛСО не только использует данные из НБДК, но и дает возможность их пополнения.

### 3.4 Поиск соседних аннотаций

*Соседние аннотации* дерева — это аннотации в НБДК, имеющие общие пары с данным деревом (формально: если пересечение множества пар, на которых построена аннотация, с множеством пар, на которых построено дерево, не является пустым).

Приведем пример запроса к НБДИ ЛСО на языке MS SQL, где требуется найти в НБДК соседние аннотации дерева с идентификатором 1:

```

select distinct
  [eq].[Id]
from [LingvoConnector].[dbo].[Equivalence] [eq]

```

```
join [LingvoConnector].[dbo].[LGF] [lg] on [lg].[Id] = [eq].[LGFOrigId]
join [LingvoConnector].[dbo].[LGF_Word] [lw] on [lw].[LGFId] = [lg].[Id]
join [LingvoText].[dbo].[WordInText] [wt] on [wt].[Id] = [lw].[WordId]
join [Structure] [s] on [s].[Id] = 1
join [LingvoText].[dbo].[Para] [p] on [p].[FragOrigId] = [wt].[FragmentId]
and [p].[ParaNum] between [s].[ParaNumStart] and [s].[ParaNumEnd]
and [p].[TranslationId] = [s].[TranslationId]
```

### 3.5 История изменений дерева

В НБДИ ЛСО фиксируется каждая операция по изменению состояний каждого дерева (создание дерева, добавление, редактирование, удаление узлов). При этом фиксируется также информация о пользователе, который произвел данное изменение. В ИСИЛСО есть возможность получить список всех изменений некоторого дерева, а также просмотреть каждое состояние и сравнить его с предыдущим (если такое имеется).

Приведем пример запроса к НБДИ ЛСО на языке MS SQL, где требуется получить список всех состояний дерева с идентификатором 10:

```
select
  [h].[Order],
  [h].[StartTime],
  [h].[EndTime],
  [u].[LastName]
from [History] [h]
join [LingvoConnector].[dbo].[Users] [u] on [u].[Id] = [h].[UserId]
where [h].[StructureId] = 10
```

Обращение к НБДК здесь необходимо потому, что в НБДИ ЛСО нет своего списка пользователей, а множества пользователей НБДИ ЛСО и НБДК совпадают.

## 4 Заключение

В статье представлены интеграционные возможности баз данных, функционирующих на основе параллельных корпусов. Такая интеграция, объединяющая несколько корпусных информационных ресурсов между собой, позволяет в методологическом отношении комбинировать наработки контрастивной лингвистики, теории и практики перевода, не упуская из виду факторы, которые могут отрицательно сказаться на валидности и надежности конечных данных.

В перспективе интеграционные возможности надкорпусных баз данных можно будет использовать при наполнении баз знаний искусственного интеллекта (в том числе в системах машинного перевода). Кроме того, благодаря некоторым своим особенностям — таким как, например, свойство фиксировать историю

изменений состояния дерева — надкорпусная база данных типа НБДИ ЛСО может стать важным инструментом в психолингвистических исследованиях речи. Для изучения процесса принятия решений при речепорождении, как правило, используют программное обеспечение, отслеживающее движение глаз или регистрирующее щелчки клавиш на клавиатуре компьютера. Здесь НБДИ ЛСО открывает новые возможности. Также история изменений состояния дерева может оказаться востребованной для теории порождения и управления знаниями, наглядно представляя всю хронологическую последовательность приложения когнитивных усилий к одному и тому же информационному объекту (и прирост нового знания, с ним связанного).

## Литература

1. Crossroads between contrastive linguistics, translation studies and machine translation: TC3 II (Translation and Multilingual Natural Language Processing 4) / Eds. O. Czulo, S. Hansen-Schirra. — Berlin: Language Science Press, 2017. 208 p.
2. *Bystrova-McIntyre T.* Looking at the overlooked: A corpora study of punctuation use in Russian and English // *Transl. Interpret. Stu.*, 2007. Vol. 2. No. 1. P. 137–162.
3. *Gast V.* On the use of translation corpora in contrastive linguistics: A case study of impersonalization in English and German // *Languages Contrast*, 2015. Vol. 15. Iss. 1. P. 4–33.
4. New directions in corpus-based translation studies / Eds. C. Fantinuoli, F. Zanettin. — Berlin: Language Science Press, 2015. 168 p.
5. *Mikhailov M., Cooper R.* Corpus linguistics for translation and contrastive studies: A guide for research. — Abingdon, New York: Routledge, 2016. 258 p.
6. Parallel corpora for contrastive and translation studies: New resources and applications / Eds. I. Doval, M. T. Sánchez Nieto. — Amsterdam, Philadelphia: John Benjamins, 2019. 310 p.
7. Translating and comparing languages: Corpus-based insights / Eds. S. Granger, M.-A. Lefer. — Louvain-la-Neuve, Belgique: Presses universitaires de Louvain, 2020. 298 p.
8. Corpora in translation and contrastive research in the digital age: Recent advances and explorations // Eds. J. Lavid-López, C. Maíz-Arévalo, J. R. Zamorano-Mansilla. — Amsterdam, Philadelphia: John Benjamins, 2021. 351 p.
9. Extending the scope of corpus-based translation studies / Eds. S. Granger, M.-A. Lefer. — London: Bloomsbury Academic, 2022. 288 p.
10. Семантика коннекторов: контрастивное исследование / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 396 с.
11. Структура коннекторов и методы ее описания / Науч. ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. 316 с.
12. *Дурново А. А., Инькова О. Ю., Попкова Н. А.* Архитектура базы данных иерархии логико-семантических отношений // *Системы и средства информатики*, 2022. Т. 32. № 1. С. 114–125. doi: 10.14357/08696527220111.
13. *Inkova O., Popkova N.* Statistical data as information source for linguistic analysis of Russian connectors // *Информатика и её применения*, 2017. Т. 11. Вып. 3. С. 123–131.

14. Национальный корпус русского языка (НКРЯ). <https://ruscorpora.ru/new>.
15. *Инькова О. Ю.* Структура коннекторов: лингвистические методы описания // Структура коннекторов и методы ее описания / Науч. ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. С. 5–47.
16. *Дурново А. А., Инькова О. Ю., Попкова Н. А.* Интерфейс пользователя Надкорпусной базы данных иерархий логико-семантических отношений // Системы и средства информатики, 2022. Т. 32. № 3. С. 103–115. doi: 10.14357/08696527220310.

*Поступила в редакцию 11.10.22*

---

---

## INTEGRATION CAPACITIES OF SUPRACORPORA DATABASES

*A. A. Durnovo<sup>1</sup>, O. Yu. Inkova<sup>1,2</sup>, and V. A. Nuriev<sup>1</sup>*

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland

**Abstract:** The paper centers on integration capacities of supracorpora databases developed at the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences. It is shown how three databases — the Supracorpora database of hierarchal logical-semantic relations (SDBH LSR), the Database of parallel texts (DBT), and the Supracorpora database of connectives (SDBC) — are integrated between themselves. The information system of hierarchal logical-semantic relations uses a specially designed database (SDBH LSR) in which annotations of logical-semantic relations are presented as trees, i. e., directed connected acyclic graphs where nodes contain data and edges depict the subordination between nodes. Along with the SDBH LSR, ISHLSR uses data from the DBT and SDBC. Such integration makes it possible to combine the methodological strengths of the informatics, contrastive and corpus linguistics, and theory and practice of translation without losing sight of the factors that may adversely affect the validity and reliability of the final data.

**Keywords:** supracorpora database; integration of databases; multilingual corpus; parallel corpus; corpus-based information resources; translation studies; contrastive linguistics; machine translation

**DOI:** 10.14357/08696527230103

### Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Информатика”) of FRC CSC RAS (Moscow) and supported by the Switzerland–Russia Research Preparation Grant “Annotation Methodology in a Supracorpora Database of Connectives” of the State Secretariat for Education, Research, and Innovation.



## References

1. Czulo, O., and S. Hansen-Schirra, eds. 2017. *Crossroads between contrastive linguistics, translation studies and machine translation: TC3 II (Translation and Multilingual Natural Language Processing 4)*. Berlin: Language Science Press. 208 p.
2. Bystrova-McIntyre, T. 2007. Looking at the overlooked: A corpora study of punctuation use in Russian and English. *Transl. Interpret. Stu.* 2(1):137–162.
3. Gast, V. 2015. On the use of translation corpora in contrastive linguistics: A case study of impersonalization in English and German. *Languages Contrast* 15(1):4–33.
4. Fantinuoli, C., and F. Zanettin, eds. 2015. *New directions in corpus-based translation studies*. Berlin: Language Science Press. 168 p.
5. Mikhailov, M., and R. Cooper. 2016. *Corpus linguistics for translation and contrastive studies: A guide for research*. Abingdon, New York: Routledge. 258 p.
6. Doval, I., and M. T. Sánchez Nieto, eds. 2019. *Parallel corpora for contrastive and translation studies: New resources and applications*. Amsterdam, Philadelphia: John Benjamins. 310 p.
7. Granger, S., and M.-A. Lefer, eds. 2020. *Translating and comparing languages: Corpus-based Insights*. Louvain-la-Neuve, Belgique: Presses universitaires de Louvain. 298 p.
8. Lavid-López, J., C. Maíz-Arévalo, and J. R. Zamorano-Mansilla, eds. 2021. *Corpora in translation and contrastive research in the digital age: Recent advances and explorations*. Amsterdam, Philadelphia: John Benjamins. 351 p.
9. Granger, S., and M.-A. Lefer, eds. 2022. *Extending the scope of corpus-based translation studies*. London: Bloomsbury Academic. 288 p.
10. Inkova, O. Yu., ed. 2018. *Semantika konnektorov: Kontrastivnoe issledovanie* [Semantics of connectives: Contrastive study]. Moscow: TORUS PRESS. 396 p.
11. Inkova, O. Yu., ed. 2019. *Struktura konnektorov i metody ee opisaniya* [Structure of connectives and methods for its description]. Moscow: TORUS PRESS. 316 p.
12. Durnovo, A. A., O. Yu. Inkova, and N. A. Popkova. 2022. Arkhitektura bazy dannykh ierarkhii logiko-semanticheskikh otnosheniy [Database of hierarchies of logical-semantic relations: Architecture]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):114–125. doi: 10.14357/08696527220111.
13. Inkova, O. Yu., and N. Popkova. 2017. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):123–131.
14. Natsional'nyy korpus russkogo yazyka [Russian national corpus]. Available at: <http://www.ruscorpora.ru> (accessed March 17, 2022).
15. Inkova, O. Yu. 2019. Struktura konnektorov: lingvisticheskie metody opisaniya [Structure of connectives: Linguistic methods of description]. *Struktura konnektorov i metody ee opisaniya* [Structure of connectives and methods of its description]. Ed. O. Yu. Inkova. Moscow: TORUSS PRESS. 5–47.
16. Durnovo, A. A., O. Yu. Inkova, and N. A. Popkova. 2022. Interfeys pol'zovatelya Nad-korpusnoy bazy dannykh ierarkhii logiko-semanticheskikh otnosheniy [Database of hierarchies of logical-semantic relations: User interface]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):103–115. doi: 10.14357/08696527220310.

*Received October 11, 2022*

## **Contributors**

**Durnovo Aleksandr A.** (b. 1949) — leading programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; duralex49@mail.ru

**Inkova Olga Yu.** (b. 1965) — Doctor of Science in philology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; faculty member, University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland; olyainkova@yandex.ru

**Nuriev Vitaly A.** (b. 1980) — Doctor of Science in philology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; nurieff.v@gmail.com

# ИСПОЛЬЗОВАНИЕ БАЗЫ ДАННЫХ МОДЕЛЕЙ СТРУКТУРНЫХ ПЕРЕВОДЧЕСКИХ ТРАНСФОРМАЦИЙ ДЛЯ ИЗВЛЕЧЕНИЯ МНОГОКОМПОНЕНТНЫХ ТЕРМИНОЛОГИЧЕСКИХ ЕДИНИЦ

*Ю. И. Бутенко*<sup>1</sup>

**Аннотация:** Раскрыты основные принципы использования базы данных моделей структурных переводческих трансформаций терминологических словосочетаний для извлечения и выравнивания многокомпонентных терминов из англо- и русскоязычных параллельных текстов. Описана формальная структура терминологических единиц, а также структурные модели многокомпонентных терминов в английском и русском языках. Дано определение структурных трансформаций в переводе, выделены основные модели структурных трансформаций англо- и русскоязычных терминов при переводе научно-технических текстов. Обосновано, что использование структурных переводческих моделей терминологических словосочетаний требует учета специфических особенностей каждого анализируемого языка. Предложен подход к извлечению и выравниванию терминологических единиц в параллельных научно-технических текстах, а каждый этап предложенного подхода проиллюстрирован примерами обработки параллельного научно-технического текста на русском и английском языках.

**Ключевые слова:** структурная модель; многокомпонентный термин; формальная структура термина; модели структурных переводческих трансформаций; русскоязычная терминология; англоязычная терминология

**DOI:** 10.14357/08696527230104

## 1 Введение

В настоящее время профессиональная межкультурная коммуникация между учеными и специалистами различных сфер профессиональной деятельности становится обыденной нормой, неотъемлемой частью которой должна быть однозначно понимаемая и непротиворечивая терминология. Значительное увеличение доли переводных научных журналов, нормативной и прочей документации свидетельствует об увеличении роли профессиональной межкультурной коммуникации и, соответственно, ведет к ужесточению требований к качеству перевода научно-технических текстов и особенно специальной терминологии. В основе упорядочения терминологии, а также поиска адекватных переводных эквивалентов лежит наличие возможности обрабатывать большие массивы неструктурированной информации, представленной текстами на естественном языке.

---

<sup>1</sup>Московский государственный технический университет им. Н. Э. Баумана, iubutenko@bmstu.ru

Несмотря на значительные успехи в обработке естественного языка, обработка научно-технических текстов сопряжена со значительными трудностями, главная из которых — формальная структура терминов [1]. Так, термин может быть однокомпонентным и состоять из ключевого слова или представлять собой терминологическую группу, в состав которой входит ключевое слово, или ядро группы, одно или несколько левых определений и одно или несколько правых или предложных определений, которые уточняют или модифицируют смысл терминологической единицы. Терминологическое словосочетание может состоять из 10–12 компонентов. При этом терминологическое словосочетание представляет собой номинативную единицу в некоторой предметной области, которая выражает одно целое понятие [2].

Согласно В. М. Лейчику, подавляющее большинство терминов имеет формальную структуру лексических единиц того естественного языка, в сфере которого они функционируют и образуются [2], т. е. многокомпонентные термины разных предметных областей в рамках одного языка имеют одни и те же способы словообразования и, соответственно, один и тот же перечень структурных моделей. Таким образом, для извлечения многокомпонентных терминов в параллельных текстах целесообразно использовать их структурные модели, установив варианты структурных трансформаций терминов при переводе с одного языка на другой.

Целью исследования ставится использование базы данных моделей структурных переводческих трансформаций многокомпонентных терминов при переводе, а на ее основе разработка подхода к извлечению и выравниванию терминов в параллельных научно-технических текстах.

## **2 Структурные модели английских и русских терминологических единиц**

Стоит отметить, что структурные модели терминологических единиц для русского и английского языков являются сравнительно хорошо изученной и описанной областью языкознания [3, 4]. К наиболее часто используемым моделям терминов в русском языке можно отнести двух- и трехкомпонентные словосочетания, образованные по следующим моделям: существительное + существительное в родительном падеже; прилагательное + существительное; прилагательное + + прилагательное + существительное; прилагательное + существительное + существительное в родительном падеже и др. Для английского языка наиболее продуктивны модели: *noun + noun*; *adjective + noun*; *noun + preposition + noun*; *noun + noun + noun* и т. д.

Структурные переводческие трансформации — это грамматические преобразования формальной структуры лексической единицы, с помощью которых можно осуществить переход от единиц на языке оригинала к единицам языка перевода в указанном смысле [5]. В таблице представлены примеры структурных пе-

## Структурные переводческие трансформации модели noun + noun

Примеры англоязычных терминов, образованных по модели noun + noun	Виды структурных переводческих трансформаций эквивалентных русскоязычных терминов	
	Модель	Пример
Stud welding	Сущ. + сущ. в родительном падеже	Приварка шпилек
Resistance welding	Прил. + прил. + сущ.	Электрическая контактная сварка
Upset welding	Прил. + сущ. + сущ. в творительном падеже	Стыковая сварка сопротивлением
Pressure welding	Сущ. + сущ. в творительном падеже	Сварка давлением
Gravity welding	Сущ. + прил. + сущ. в творительном падеже	Сварка наклонным электродом
Dielectric welding	Прил. + сущ. + сущ. в родительном падеже	Высокочастотная сварка пластмасс

реводческих трансформаций англоязычных терминов, образованных по модели noun + noun, а также структурные переводческие трансформации их эквивалентов в русском языке.

### 3 Применение базы данных моделей структурных переводческих трансформаций для извлечения и выравнивания многокомпонентных терминологических единиц

В настоящее время выравнивание терминологических единиц в параллельных текстах обычно осуществляется в два этапа. Сначала выделяют терминологические единицы на каждом языке отдельно. Затем один из одноязычных списков терминов-кандидатов интерпретируется как язык источника, и для каждого термина-кандидата на языке-источнике предлагаются потенциально эквивалентные термины в списке терминов-кандидатов на языке перевода [6].

Предлагаемый подход к автоматическому извлечению и выравниванию терминов на основе базы данных моделей структурных переводческих трансформаций терминологических словосочетаний состоит из 5 основных этапов. В качестве примера рассмотрим извлечение терминов для следующего фрагмента текста и его перевода:

Отсутствие аккомодации свойственно не только явлению переноса количества движения, но и, как показывают экспериментальные исследования, процессу обмена энергией между падающими молекулами и стенкой.

The absence of accommodation is characteristic not only of the phenomenon of momentum transfer, but also, as experimental studies show, the process of energy exchange between incident molecules and the wall.

На первом этапе проводится автоматический морфологический анализ текста, т. е. каждому слову приписываются его грамматические характеристики [6]:

Отсутствие	СУЩ, неод, ср ед, им	аккомодации	СУЩ, неод, жр ед, рд	свойственно	Н	не	ЧАСТ	The	determiner	absence	noun	of	prep
только	ЧАСТ	явлению	СУЩ, неод, ср ед, дл	явлению	СУЩ, неод, ср ед, дл			accommodation	noun	is	verb	charac-	
переноса	СУЩ, неод, ср ед, рд	количества	СУЩ, неод, ср ед, рд	движения	СУЩ, неод, ср ед, рд			teristic	adj	not	adverb	only	adverb
ЭНР	СОЮЗ	ЭНР	СОЮЗ	ЭНР	СОЮЗ	ЭНР	СОЮЗ	of	prep	the	determiner	phenome-	
казывают	ГЛ, несов, неперех, наст, действ, мн, тв	показывают	ГЛ, несов, неперех, наст, действ, мн, тв	исследования	СУЩ, неод, ср ед, рд	ЭНР	процессу	СУЩ, неод, ср ед, дл	non	noun	of	prep	momentum
исследования	СУЩ, неод, ср ед, рд	обмена	СУЩ, неод, ср ед, рд	энергией	СУЩ, неод, жр ед, тв	между	ПР	пада-	transfer	noun	,	but	conj
молекулами	СУЩ, неод, жр мн, тв	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ющими	as	prep	experimental	studies	noun
ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	show	verb	,	the	determiner
ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	of	prep	energy	noun	exchange
ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	between	prep	incident	adj	mole-
ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	cules	noun	and	conj	the
ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	ЭНР	wall	noun	.		

На втором этапе из размеченного текста необходимо убрать части речи, которые не входят в состав терминологических словосочетаний, такие как глаголы, местоимения, междометия и т. д., в том числе и знаки препинания [7]. После удаления любого из элементов, не входящих в состав терминологического словосочетания, следующая цепочка слов начинается с новой строки. В результате обработки рассматриваемого фрагмента текста на втором этапе получаем список терминов-кандидатов:

Отсутствие	СУЩ, неод, ср ед, им	аккомодации	СУЩ, неод, жр ед, рд	свойственно	Н	явлению	СУЩ, неод, ср ед, дл	переноса	СУЩ, неод, ср ед, рд	количества	СУЩ, неод, ср ед, рд	движения	СУЩ, неод, ср ед, рд	исследования	СУЩ, неод, ср ед, рд	процессу	СУЩ, неод, ср ед, дл	обмена	СУЩ, неод, ср ед, рд	энергией	СУЩ, неод, жр ед, тв	между	ПР	падающими	СУЩ, неод, жр мн, тв	молекулами	СУЩ, неод, жр мн, тв	стенкой	СУЩ, неод, жр ед, тв	ЭНР

На третьем этапе выполняется проверка ряда лингвистических условий [8]. Например, терминологическое словосочетание как в английском, так и русском языке не может начинаться с предлога: of<sup>prep</sup> the<sup>determiner</sup> phenomenon<sup>noun</sup> of<sup>prep</sup> momentum<sup>noun</sup> transfer<sup>noun</sup>, а также состоять из одного элемента, кроме существительного, например наречия: also<sup>adverb</sup>. В первом случае предлог исключается из словосочетания, а во втором слово удаляется из списка терминов-кандидатов. В английском языке также опускаются артикли. Стоит отметить,

что использование методов автоматического извлечения многокомпонентных терминов еще усложнено тем, что часто многокомпонентные термины сочетаются с общей лексикой, а их синтаксическая конструкция совпадает со структурной моделью. Например, прил. + прил. + сущ. — *интеллектуальная информационная технология* (термин) и *современная информационная технология* (общеупотребительное слово + термин). В таком случае необходимо сформировать список общеупотребительных слов с оценочной или внепредметной семантикой (*современный, передовой, новый и т. п.*), чтобы не учитывать их при извлечении терминов.

На четвертом этапе проводится сравнение полученных словосочетаний со структурными моделями терминологических словосочетаний. При этом при сравнении полученных на предыдущем этапе фрагментов текста со структурными моделями терминов ядерный элемент термина может иметь любой падеж и число, а остальные элементы строго соответствовать структурной модели. В русском языке ядерным элементом всегда служит имя существительное, вступающее в словоизменительную парадигму в зависимости от языкового окружения в предложении, прилагательное наследует грамматические характеристики имени существительного, перед которым стоит, а все остальные элементы обладают неизменными грамматическими характеристиками, например: **сущ.** + прил. + сущ. в творительном падеже — **сварка/сварки/сваркой** *наклонным электродом*. В английском языке ядерным элементом служит последнее имя существительное в цепочке или последнее существительное перед предлогом, например: noun + **noun** — information **technology** и noun + preposition + noun — **theory** of translation.

На пятом этапе ядерные элементы каждого терминологического словосочетания приводятся к форме именительного падежа единственного числа, затем попарно сопоставляются с моделями переводческих трансформаций из базы данных и проверяются на наличие в терминологическом словаре. Оставшиеся словосочетания передаются на ручную обработку:

аккомодации <sup>СУЩ,неод,жр ед,рд</sup>	явлению <sup>СУЩ,неод,сред,дт</sup>	accommodation <sup>noun</sup>
переноса <sup>СУЩ,неод,мр ед,рд</sup>	количества <sup>СУЩ,неод,сред,рд</sup>	phenomenon <sup>noun</sup> of <sup>prep</sup>
движения <sup>СУЩ,неод,сред,рд</sup>	эксперименталь- ные <sup>ПРИЛ неод,мн,вн</sup>	momentum <sup>noun</sup> transfer <sup>noun</sup>
процессу <sup>СУЩ,неод,мр ед,дт</sup>	исследования <sup>СУЩ,неод,сред,рд</sup>	experimental studies <sup>noun</sup>
энергией <sup>СУЩ,неод,жр ед,тв</sup>	обмена <sup>СУЩ,неод,мр ед,рд</sup>	process <sup>noun</sup> of <sup>prep</sup>
ми <sup>ПРИЧ,несов,неперех,наст,действмн,тв</sup>	между <sup>ПР</sup>	energy <sup>noun</sup> exchange <sup>noun</sup>
ми <sup>СУЩ,неод,жр мн,тв</sup>	стенкой <sup>СУЩ,неод,жр ед,тв</sup>	between <sup>prep</sup> incident <sup>adj</sup>
		molecules <sup>noun</sup> wall <sup>noun</sup>

Такой подход позволяет значительно сократить объем ручного труда при извлечении и выравнивании терминологических единиц, а также может быть расширен на другие языки [9]. При этом он позволяет находить неологизмы, которые еще не имеют переводного эквивалента в языке перевода, а переводятся описательным способом, например: *squirt welding* — *полуавтоматическая дуговая сварка под флюсом с подачей флюса из бункера, укрепленного на дер-*

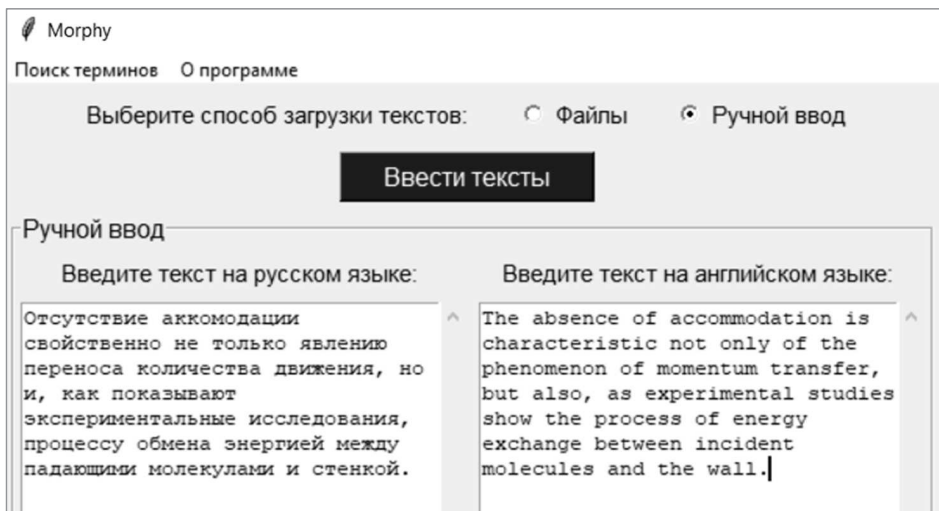


Рис. 1 Интерфейс системы извлечения многокомпонентных терминов

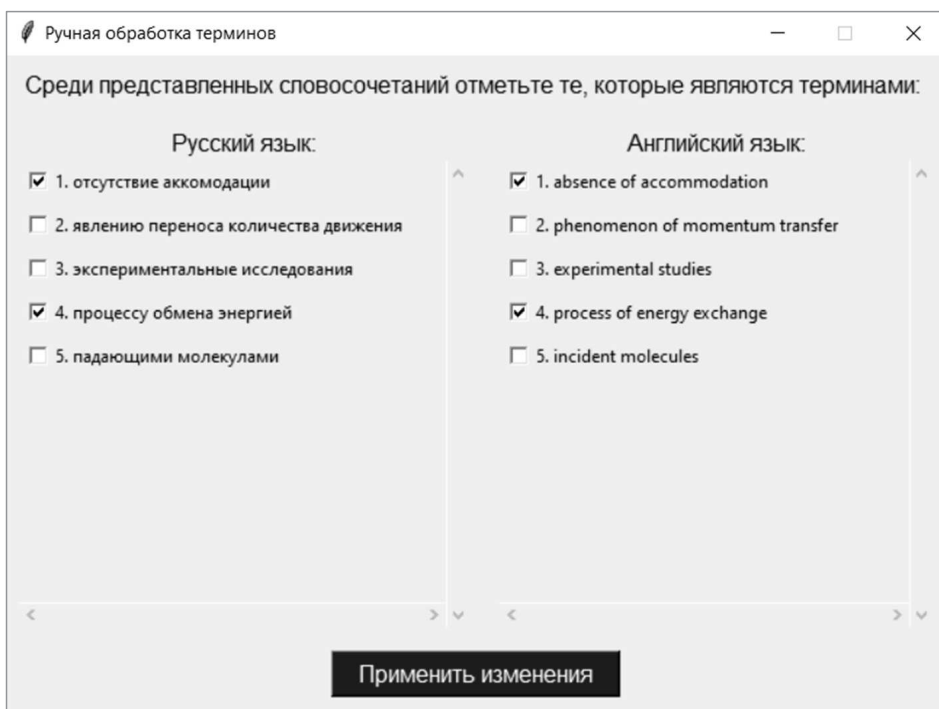


Рис. 2 Выбор терминов и их переводных эквивалентов



*жатели*, что является крайне важным при исследовании процессов образования и перевода терминов.

На основе предложенного метода разработано приложение (рис. 1), написанное на языке Python с использованием библиотек Tkinter, NLTK, rumporphy2, os, sys. В результате работы приложения отобранные термины выводятся на экран, после чего их можно классифицировать и сохранять в базу данных (рис. 2).

Использование приложения позволяет в несколько раз сократить время на поиск терминов и их эквивалентов в параллельных текстах. Так, при обработке научно-технической статьи по космонавтике объемом 2 645 слов в русской версии и 3 041 слово в переводе на английский приложение выделило 412 пар термин-кандидат и его переводной эквивалент. Число выделенных приложением пар будет уменьшаться по мере наполнения базы данных терминами и их переводными эквивалентами. При этом эффективность работы приложения также зависит от тематики текста (технические тексты могут содержать номенклатурные наименования или литеральные термины, включающие символы других алфавитов или обозначения химических элементов).

#### 4 Заключение

В работе предложен подход к извлечению и выравниванию терминов в параллельных научно-технических текстах на основе базы данных моделей структурных переводческих трансформаций многокомпонентных терминов при переводе. В ходе исследования установлено, что многокомпонентные термины разных предметных областей в рамках одного языка имеют одни и те же способы словообразования и, соответственно, один и тот же перечень структурных моделей, что позволяет использовать их при извлечении многокомпонентных терминов. К наиболее часто используемым моделям терминов в русском и английском языках относят двух- и трехкомпонентные словосочетания. Модели структурных переводческих трансформаций терминов отражают структуру термина в языке-источнике и его наиболее вероятные структуры в языке перевода. Предлагаемый подход к автоматическому извлечению и выравниванию терминов на основе базы данных моделей структурных переводческих трансформаций терминологических словосочетаний состоит из следующих основных этапов:

- (1) морфологический анализ текста;
- (2) отсеивание частей речи, которые не входят в состав терминологических словосочетаний;
- (3) проверка лингвистических ограничений на качественный состав терминологической единицы;
- (4) сравнение словосочетаний со структурными моделями терминологических словосочетаний;
- (5) проверка по терминологическому словарю или передача на ручную обработку.

## Литература

1. *Butenko Yu. I., Garazha V. V.* BMSTU corpus of scientific and technical texts: Conceptual framework // *Applied Linguistics Research J.*, 2021. Vol. 5. Iss. 3. P. 76–81. doi: 10.14744/alrj.2021.15579.
2. *Лейчик В. М.* Исходные понятия, основные положения, определения современного терминоведения и терминографии // *Вестник Харьковского политехнического ун-та*, 1994. № 1. С. 147–180.
3. *Гринева-Гринева С. В., Сорокина Э. А.* Опыт описания формальной структуры термина (на материале английской терминологии лексикологии) // *Вестник Московского государственного областного ун-та. Сер. Лингвистика*, 2020. № 5. С. 74–85. doi: 10.18384/2310-712X-2020-5-74-85.
4. *Бутенко Ю. И., Николаева Н. С., Карцева Е. Ю.* Структурные модели англоязычных терминов для автоматической обработки корпусов научно-технических текстов // *Вестник Российского университета дружбы народов. Сер. Теория языка. Семиотика. Семантика*, 2022. Т. 13. № 1. С. 80–95. doi: 10.22363/2313-2299-2022-13-1-80-95.
5. *Лейчик В. М., Шелов С. Д.* Лингвистические проблемы терминологии и научно-технический перевод. — М.: Всесоюз. центр переводов науч.-техн. лит. и документации, 1990. 78 с.
6. *Simon N. I., Kešelj V.* Automatic term extraction in technical domain using part-of-speech and common-word features // *ACM Symposium on Document Engineering Proceedings*. — New York, NY, USA: ACM, 2018. Art. 51. 4 p. doi: 10.1145/3209280.3229100.
7. *Сидорова Е. А.* Комплексный подход к исследованию лексических характеристик текста // *Вестник СибГУТИ*, 2019. № 3. С. 80–88.
8. *Бизюкова Н. Ю., Тарасова О. А., Рудик А. В., Филимонов Д. А., Поройков В. В.* Автоматическое распознавание названий химических соединений в текстах научных публикаций // *Научно-техническая информация. Сер. 2: Информационные процессы и системы*, 2020. № 11. С. 36–46. doi: 10.36535/0548-0027-2020-11-5. EDN: EPSCAL.
9. *Vecerro F. B.* Phraseological variations in medical-pharmaceutical terminology and its applications for English and German into Spanish translations // *SciMedicine J.*, 2020. Vol. 2. Iss. 1. P. 22–29.

*Поступила в редакцию 06.09.22*

---

---

## USING A DATABASE OF STRUCTURAL TRANSFORMATIONS TO EXTRACT MULTICOMPONENT TERMINOLOGICAL UNITS

*Yu. I. Butenko*

N. E. Bauman Moscow State Technical University, 5-1 Baumanskaya 2nd Str.,  
Moscow 105005, Russian Federation

**Abstract:** The paper reveals the basic principles of using the database of structural models of terminological phrases for the extraction and alignment of

English and Russian multicomponent terms. The formal structure of terminological units and structural models of multicomponent terms in English and Russian are described. The definition of structural transformations in translation is given and the main models of structural transformations of English- and Russian-speaking terms in the translation of scientific and technical texts are highlighted. It is substantiated that the use of structural models of terminological word combinations requires taking into account the specific features of each language under analysis. An approach to the extraction and alignment of terminological units in parallel scientific and technical texts is proposed. It consists of 5 main stages. The approach to the extraction and alignment of terminological units is illustrated by examples of processing the parallel scientific and technical text in Russian and English.

**Keywords:** structural model; multicomponent term; formal structure of the term; structural models of terminological phrases in translation; Russian-language terminology; English-language terminology

**DOI:** 10.14357/08696527230104

## References

1. Butenko, Iu. I., and V. Garazha. 2021. BMSTU corpus of scientific and technical texts: Conceptual framework. *Applied Linguistics Research J.* 5(3):76–81. doi: 10.14744/alrj.2021.15579.
2. Lejchik, V. M. 1994. Iskhodnye ponyatiya, osnovnye polozheniya, opredeleniya sovremennogo terminovedeniya i terminografii [Linguistic problems of terminology and scientific and technical translation]. *Bulletin of the Kharkiv Polytechnic Institute* 1:147–180.
3. Grinev-Grinevich, S. V., and E. A. Sorokina. 2020. Opyt opisaniya formal'noy struktury termina (na materiale angliyskoy terminologii leksikologii) [Describing the formal structure of a term (based on the English terminology of lexicology)]. *Bulletin of the Moscow Region State University. Ser. Linguistics* 5:74–85. doi: 10.18384/2310-712X-2020-5-74-85.
4. Butenko, Iu. I., N. S. Nikolaeva, and E. Yu. Kartseva. 2022. Strukturnye modeli anglo-yazychnykh terminov dlya avtomaticheskoy obrabotki korpusov nauchno-tekhnicheskikh tekstov [Structural models of English terms of automated processing of scientific and technical texts corpora]. *RUDN J. Language Studies Semiotics Semantics* 13(1):80–95. doi: 10.22363/2313-2299-2022-13-1-80-95.
5. Lejchik, V. M., and S. D. Shelov. 1990. *Lingvisticheskie problemy terminologii i nauchno-tekhnicheskiiy perevod* [Linguistic problems of terminology and scientific and technical translation]. Moscow: Vsesoyuz. tsentr perevodov nauch.-tekhn. lit. i dokumentatsii. 78 p.
6. Simon, N. I., and V. Kešelj. 2018. Automatic term extraction in technical domain using part-of-speech and common-word features. *ACM Symposium on Document Engineering Proceedings*. New York, NY: ACM. 51. 4 p. doi: 10.1145/3209280.3229100.
7. Sidorova, E. A. 2019. Kompleksnyy podkhod k issledovaniyu leksicheskikh kharakteristik teksta [The integrated approach to text lexical characteristics study]. *Vestnik SiBGUTI* [The Herald of the Siberian State University of Telecommunications and Informatics] 3:80–88.

8. Biziukova, N. Y., O. A. Tarasova, A. V. Rudik, *et al.* 2020. Automatic recognition of chemical entity mentions in texts of scientific publications. *Automatic Documentation Mathematical Linguistics* 54(6):306–315. doi: 10.3103/S0005105520060023.
9. Becerro, F. B. 2020. Phraseological variations in medical-pharmaceutical terminology and its applications for English and German into Spanish translations. *SciMedicine J.* 2(1):22–29.

*Received September 6, 2022*

### **Contributor**

**Butenko Yuliya I.** (b. 1987) — Candidate of Science (PhD) in technology, associate professor, N. E. Bauman Moscow State Technical University, 5-1 Baumanskaya 2nd Str., Moscow 105005, Russian Federation; iubutenko@bmstu.ru

## СЕМАНТИЧЕСКИЕ ИНТЕРПРЕТАЦИИ ВЫСОКИХ НОРМАЛЬНЫХ ФОРМ ОТНОШЕНИЙ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

*В. А. Иванов<sup>1</sup>, М. Ю. Коньшев<sup>2</sup>, С. В. Смирнов<sup>3</sup>, О. В. Тараканов<sup>4</sup>,  
В. О. Тараканова<sup>5</sup>, С. В. Усовик<sup>6</sup>*

**Аннотация:** Приведены результаты семантического моделирования процессов устранения избыточности и защиты отношения реляционной базы данных (БД) от аномалий обновления за счет совершенствования процесса его нормализации. Установлены семантические эквиваленты требований нормальных форм реляционного отношения, повышающие результативность алгоритма нормализации. Сформулирована и доказана теорема, гарантирующая однозначную детерминацию функциональной зависимости (ФЗ) атрибута от потенциального ключа. Сформулированы уточнения алгоритма нормализации реляционного отношения. Приведены способы применения предложенных интерпретаций. Показаны практически применимые процедуры, снимающие противоречия проектирования рациональной структуры БД. Основные выводы сформулированы в отношении фактографических БД. Предложенные механизмы не зависят от используемой системы управления БД.

**Ключевые слова:** реляционная база данных; нормализация отношения; потенциальный ключ; неключевой атрибут; функциональная зависимость; многозначная зависимость

**DOI:** 10.14357/08696527230105

### 1 Введение

Высокая согласованность, целостность и непротиворечивость фактографических реляционных БД служит весомым аргументом в пользу их широкого использования. При этом ограничивающим фактором становятся низкие показатели интенсивности результатных свойств, обусловленные их неквалифицированным проектированием. В отрасли БД наблюдается нестабильность подходов к вопросу их эффективного проектирования [1]. Преобладают мнения о том, что огрехи проектирования можно скомпенсировать обработчиками на языке программирования высокого уровня, избыточность контролировать с помощью специального

---

<sup>1</sup> Российский технологический университет — МИРЭА, iva.mac@mail.ru

<sup>2</sup> Финансовый университет при Правительстве РФ, mykonyshev@fa.ru

<sup>3</sup> Российский технологический университет — МИРЭА, smirnov\_s@mirea.ru

<sup>4</sup> Российский технологический университет — МИРЭА, ole\_g66@list.ru

<sup>5</sup> Федеральная служба охраны РФ, talero@list.ru

<sup>6</sup> Российский технологический университет — МИРЭА, usovik@mirea.ru

программного кода, что порождает ложную уверенность в незначимости обеспечения целостности, согласованности и непротиворечивости.

Имеется множество определений целостности, например: целостность — это соответствие имеющейся в БД информации ее внутренней логике, структуре и всем явно заданным правилам [2]. Или целостность — это соответствие значений всех данных определенному непротиворечивому набору правил [3], а также целостность означает, что данные полны, условие того, что данные не были изменены при выполнении любой операции над ними (кроме целенаправленного изменения), будь то передача, хранение или представление [4]. Целостность данных — их свойство, предполагающее, что точность и согласованность сохраняются независимо от внесенных изменений [5].

Важное значение нормализации в обеспечении целостности данных и целостности БД обуславливает актуальность приведения данной процедуры к наивысшей степени эффективности ее реализации. Под эффективностью нормализации следует понимать ее высокую результативность и экономичность. Требуемый уровень данных показателей в настоящее время достигается вручную, точным исполнением разработчиком предписаний

## **2 Постановка задачи исследования**

Основной парадигмой нормализации считается избавление отношений реляционной БД от нежелательных зависимостей между атрибутами. Но у этой стройной теории имеется ряд недостатков. По мнению авторов, главный из них — слабая семантическая интерпретируемость основных положений и закономерностей теории. Это приводит к ограничению в применении, побуждает специалистов искать обходные пути, подбирать суррогатные решения для компенсации нежелательных эффектов некачественно спроектированной БД.

Семантическая интерпретация ФЗ одного атрибута (зависимой части) от другого (детерминанта) через понятие математической функции [6–8] существенно повышает понятность данной части теории нормализации. Требуются аналогичные интерпретации других постулатов. Решение данной задачи, предположительно, повысит понятность механизмов нормализации, а значит, их востребованность и массовую воспроизводимость. Отсюда следует задача изыскания семантических интерпретаций основных положений теории нормализации отношений реляционной БД.

Известно, что алгоритм нормализации в пределе не сходится [9, 10]. Это вполне объяснимо с позиции слабой формализации цели данного целенаправленного процесса. Поэтому в подавляющем числе научных источников ведется речь об обобщенных или приближенных алгоритмах нормализации.

Обобщенные алгоритмы, как правило, приводят к инвариантным результатам. Данное положение справедливо и в отношении алгоритма нормализации. Инвариантность траекторий решения задачи формирования свободной от неже-

лательных зависимостей структуры БД обуславливает существование множества пригодных результатов.

Значительный вклад в снижение востребованности процедур нормализации вносит такое явление, как «парадокс нормализации». Его сущность, заключающаяся в росте числа соединяемых отношений с повышением уровня нормализованности БД, приводящая к повышению латентности запросной подсистемы, вызывает зачастую необоснованные опасения по поводу общего быстродействия информационной системы [11, 12].

Представляется, что корни описанной проблемы уходят в недостаточное понимание семантики процедуры нормализации. Известны различные трактовки нормальных форм, обеспечивающие понимание сути выполняемых преобразований [13]. При этом авторы формулируют семантические конструкции относительно примерных предметных областей.

Сущность недостаточности семантических интерпретаций нормальных форм состоит в отсутствии универсальных механизмов реабстрагирования основных конструктивов процедуры нормализации. Данная парадигма опирается на положения методологии «сущность–связь» П. Чена [14]. Процедура абстрагирования транслируется на связи и агрегаты сущностей. Это означает, что каждый произвольный фрагмент инфологической модели может быть абстрагирован детерминированным объектом. Порождаемые объекты наделены семантикой, что обеспечивает однозначность перехода от физических объектов к абстракциям.

Очевидно, что существование прямого абстрагирования обуславливает существование обратного процесса. Поиск семантической составляющей реабстрагирования — это и есть предмет исследования, описанного в данной работе. Такая постановка обратна задаче, сформулированной в известном труде М. Ш. Цаленко [6], следовательно, может считаться корректной.

Ожидается, что семантические интерпретации высоких нормальных форм реляционного отношения обеспечат сходимость, по крайней мере при соблюдении несложных условий, алгоритма нормализации. В свою очередь сходящийся алгоритм нормализации породит качественную структуру БД, свободную от аномалий и избыточности.

В теории БД многие методы и способы хорошо иллюстрируются примерами, которые позволяют достаточно легко понять сущность ФЗ, многозначную зависимость (МЗ), зависимость соединения (ЗС) и свойства зависимостей. Следовательно, повышается качество проектирования БД. Очевидно, что с помощью контрпримеров можно скомпрометировать объективные методы. По мнению авторов, способ толкования методов нормализации отношений на основе семантики требований к высоким нормальным формам имеет значительный потенциал как в вопросах логического доказательства, так и в исследованиях по совершенствованию требований.

В отрасли создания фактографических БД известно противоречие, заключающееся в том, что для получения качественной структуры (схемы) необходимо выполнить значительное число операций преобразования исходного описания

объектов учета. В работах [14, 15] сформулированы положения, обуславливающие принадлежность алгоритма проектирования оптимальной структуры отношений БД к классу вычислительно сложных, а значит, стремление к получению качественной структуры БД существенно ограничивается имеющимися для проектирования вычислительными ресурсами. Данное положение обуславливает актуальность изыскания неформальных способов формирования квазиоптимальных схем фактографических реляционных БД, составляющего предмет исследования.

Цель работы состоит в устранении ограничения алгоритма проектирования оптимальной структуры (схемы) фактографической реляционной БД по допустимой ресурсоемкости (вычислительный ресурс) путем разработки процедур семантического моделирования ФЗ, составляющих объект преобразования. Задачи исследования сводятся к семантической интерпретации необходимых и достаточных условий приведения отношений проектируемой реляционной БД к нормальным формам. Здесь подлежит уточнению понятие «функциональная зависимость», изыскание приемов устранения МЗ и обеспечения корректной ЗС.

### 3 Анализ существующих решений

В известной постановке задача синтеза неоптимальной схемы БД сформулирована в следующем виде [16, 17]. Пусть задана схема  $S = (R; F)$ , где  $R$  — множество атрибутов универсального отношения;  $F$  — множество ФЗ атрибутов из  $R$ . Необходимо получить схему  $T$  как множество подсхем (схем отношений) вида  $S_i = (R_i; K_i)$ , где  $R_i$  — множество атрибутов подсхемы;  $K_i = \{X_{i1}, \dots, X_{iq}\}$  — множество ключей подсхемы ( $X_{ij} \subseteq R_i, i = \overline{1, p}, j = \overline{1, q}$ ). При этом схема  $T$  должна удовлетворять следующим требованиям.

1. Декомпозиция  $\{R_1, \dots, R_p\}$  обладает свойством естественного соединения отношений без потерь (выполняется требование пятой нормальной формы (5НФ)).
2. Обеспечивается сохранение множества всех ФЗ из замыкания  $F^+$  (из объединения всех ФЗ подсхем логически следуют все зависимости, принадлежащие  $F$ ).
3. Все подсхемы  $S_i$  приведены к третьей нормальной форме — ЗНФ (не содержат транзитивных ФЗ неключевых атрибутов от всех потенциальных ключей).
4. Число подсхем в схеме минимально. Это означает, что не существует схемы, удовлетворяющей требованиям 1–3 и содержащей менее  $p$  подсхем.

Свойства обеспечения целостности данных на уровне одного отношения и множества ФЗ, декларированные пп. 1 и 2, существенно значимы, так как позволяют восстановить исходную схему  $S$  по декомпозиции  $T$ . Соответствие подсхем  $S_i$



ЗНФ (п. 3) позволяет избежать значительной части аномалий обновления кортежей БД. Требование п. 4 позволяет обеспечить неизбыточность БД. Оптимальная структура (схема) БД при выполнении требований 1–4 должна иметь возможно меньшее общее число атрибутов в подсхемах и минимальное множество ключевых атрибутов (обеспечить минимальность первичных ключей).

Известное решение задачи синтеза (неоптимальной) схемы, удовлетворяющей требованиям 1–4, сформулировано Бернштейном в виде алгоритма [16], который отличается сложной машинной реализацией. Уменьшение вычислительной сложности алгоритма может быть достигнуто путем уточнения и детализации его отдельных операций. Семантический анализ шагов алгоритма открывает возможности исключения отдельных работ и процедур общей последовательности. При этом возможна потеря общей алгоритмичности в отличие от решений, предложенных в [17]. Здесь требуется достижение компромисса между необходимостью детерминированности шагов алгоритма и обоснованностью их выполнения на семантическом уровне процедуры проектирования. Выигрыш может оказаться незначительным, а может существенно сократить объем рутинных операций, потребляющих ресурсы вычислительной системы.

#### 4 Ход исследования

Известные формулировки первой нормальной формы (1НФ) [10, 13] определяют достижение атомарности атрибутов для приведения отношения к ней. Последующие методики нормализации опираются на понятие ключа, относительно которого исследуются неключевые атрибуты. При этом не детерминировано требование и не описана процедура формирования полного списка нетривиальных ключей. Следует особо отметить, что речь ведется о потенциальных ключах, а выбор первичного проводится при приведении отношения к доменно-ключевой нормальной форме (ДКНФ), внешние формируются в процессе связывания отношений (процедура миграции ключа отношения).

Для определения потенциального ключа предлагается следующая

**Теорема.** *Если исследуется атрибут  $X$  отношения  $R$  на предмет детерминации ФЗ от него атрибута  $Y$ , то справедливо следующее утверждение: при изменении значения атрибута  $X$  значение атрибута  $Y$  может измениться или остаться прежним; если изменить значение атрибута  $Y$ , то при ФЗ его от детерминанта  $X$  значение последнего обязательно изменится, иначе ФЗ отсутствует:*

$$\left. \begin{aligned} (X = X^*)|_{Y:=Y^*}, X \rightarrow Y &\text{ — истина;} \\ (X \neq X^*)|_{Y:=Y^*}, X \rightarrow Y &\text{ — ложь.} \end{aligned} \right\} \quad (1)$$

**Доказательство:** изменение значения потенциального ключа возможно только при переходе к иному объекту учета, следовательно, смене фокуса курсора

в таблице БД, что обязательно по свойству ограничения целостности реляционной модели данных, запрещающему изменять или оставлять без значения ключ. Значения неключевых атрибутов для разных кортежей могут совпадать или отличаться. Изменение значения неключевого атрибута  $Y := Y^*$  при фиксированном во времени состоянии БД (состоянии предметной области) возможно только при переводе курсора на иной кортеж (установка фокуса на ином объекте учета). Неизменность значения детерминанта в данных условиях возможна только в случае, если он не является таковым, тогда ФЗ между атрибутами  $X$  и  $Y$  отсутствует. Теорема доказана.

Теорема справедлива и в случае составного детерминанта: в доказательстве следует допустить, что совокупность атрибутов потенциального ключа — это объединение их значений. Тогда на основании дистрибутивного закона алгебры Буля для нетривиальных потенциальных ключей справедливо утверждение теоремы (1) о единственности значения неключевого атрибута для текущего значения составного потенциального ключа.

Следствием теоремы является утверждение о том, что вариации неключевых атрибутов служат надежным инструментом отыскания потенциальных ключей. Атрибуты, которые не изменяют своего значения при вариации атрибута  $Y$  должны быть исследованы на принадлежность к множеству детерминантов отношения, иначе их также следует считать неключевыми.

Теорема Хита гарантирует, что при проецировании в подотношения с общим атрибутом гарантируется операция натурального соединения, обеспечивающая восстановление исходного отношения. При этом не специфицируются условия такой проекции. Очевидно, что если имеется отношение  $r$  со схемой  $R(A, B, C)$ , то доступны три варианта проецирования:

$$\begin{aligned} R_1(A, B) \text{ JOIN } R_2(A, C); \\ R_3(B, A) \text{ JOIN } R_4(B, C); \\ R_5(C, A) \text{ JOIN } R_6(C, B). \end{aligned}$$

Но  $R_1 \equiv R_3$ ,  $R_2 \equiv R_5$  и  $R_4 \equiv R_6$ . Отсюда следует, что исходное отношение может быть восстановлено в результате операции  $R_3(B, A) \text{ JOIN } R_2(A, C)$ , или  $R_2(A, C) \text{ JOIN } R_6(C, B)$ , или  $R_5(C, A) \text{ JOIN } R_4(B, C)$ . Данное утверждение ошибочно, так как не учитывает ФЗ отношений  $R_i$ .

Функциональные зависимости подотношений должны были быть установлены на этапах приведения к низшим (до нормальной формы Бойса–Кодда — НФБК включительно) нормальным формам. Если это требование было выполнено, то можно утверждать, что существует единственная совокупность проекций отношения  $r$ , свободная от эффекта соединения с потерей целостности (достигнута декомпозиция с соединением без потерь).

Методика информационного моделирования П. Чена (P. Chen) «сущность–связь» предполагает, что все атрибуты отношения — это абстракции значимых свойств объектов учета [14]. Если это так, то все атрибуты отношения тривиально

и нетривиально функционально зависят только от потенциальных ключей; следовательно, каждое из них абстрагирует класс однородных предметов (явлений, процессов) реального мира, и ничто другое.

Справедливо, что отношение считается приведенным к четвертой нормальной форме (4НФ) тогда и только тогда, когда оно уже приведено к НФБК и не существует ни одного атрибута, абстрагирующего значимое свойство некоего другого объекта учета. Иными словами, обеспечена однородность класса предметов (явлений, процессов) реального мира, абстрагированного отношением. Это отвечает требованию дополнения как одного из правил вывода Армстронга [18, 19].

Многозначная зависимость в отношении  $r$  [18] существует тогда и только тогда, когда множество значений атрибута  $B$ , соответствующее заданной паре  $[a : A; c : C]$  отношения  $r$ , зависит от  $a$  и не зависит от  $c$ . Известные семантические интерпретации МЗ [13] сводятся к подбору совокупностей кортежей, где наблюдается зависимость между атрибутами на множестве (два и более) хранимых значений. При этом утверждается, что ситуация, когда исследуемые атрибуты состоят в ФЗ друг от друга, не становится источником аномалий, а следовательно, допустима с точки зрения обеспечения целостности отношения.

Если отношение  $r$  уже приведено к НФБК, то это означает, что все его неключевые атрибуты состоят в полной нетранзитивной ФЗ от всех потенциальных ключей. Тогда рассмотрению подлежат только возможные МЗ между неключевыми атрибутами.

Примеры МЗ, приведенные в работах по нормализации отношений [10, 13], для обеспечения методической составляющей подобраны искусственно и в большинстве случаев не удовлетворяют требованиям к НФБК. Если скрупулезно провести проектирование отношений по методике «сущность–связь» П. Чена, затем выполнить обязательные проекции в соответствии с требованиями первой, второй, третьей нормальных форм и НФБК, то подобрать пример, где отсутствует ФЗ (на единичных значениях) между неключевыми атрибутами при проявлении МЗ (на множестве значений) практически невозможно. По крайней мере, авторам это до сих пор не удалось.

Изыскание примеров проявления МЗ в высоконормализованных отношениях привело к появлению гипотезы о том, что данное явление на практике невозможно. Предположительно, возникновение МЗ в высоконормализованном отношении — это следствие ошибки выполнения проецирования в ходе приведения отношения к НФБК в смысле последовательного приведения к первой, второй и третьей нормальной форме. По мнению авторов, наиболее часто такие ошибки встречаются в случае необеспечения полноты множества потенциальных ключей или ошибок в определении ФЗ неключевых атрибутов от потенциальных ключей. На основании приведенных положений предлагается гипотеза, в соответствии с которой можно считать, что возникновение МЗ между неключевыми атрибутами отношения, не состоящими в ФЗ друг с другом, возможно только в случае, когда отношение абстрагирует более чем один однородный объект учета предметной области.

Предлагается семантическая интерпретация 4НФ отношения как совокупности атрибутов, удовлетворяющих требованиям НФБК и являющихся абстракциями исключительно значимых свойств моделируемого в соответствии с методологией «сущность–связь» однородного объекта учета.

Практически ценный вывод из теоремы Фейджина [13] определяет условие приведения отношения к 4НФ: все МЗ, имеющиеся в отношении, считаются доказанными (проверенными) ФЗ. С другой стороны, если МЗ не является функциональной, то такое отношение не является приведенным к 4НФ.

Другое утверждение, часто приводимое в научных и методических публикациях, декларирует свойство «двойственности» МЗ, когда существует пара атрибутов — зависимых частей от одного и того же детерминанта —  $A \rightarrow B|C$ , выводимое как  $(A \rightarrow B) \Leftrightarrow (A \rightarrow C)$  [10, 13, 18]. Приводимые в доказательство примеры содержат отношения, состоящие из атрибутов, служащих абстракциями свойств разнородных объектов учета. Для исследования явления МЗ между абстракциями свойств однородных объектов учета пригодный пример подобрать не удалось. Этот факт порождает предположение, что приведение отношения к состоянию, когда им абстрагируется класс однородных объектов учета, обеспечивает выполнение условий 4НФ.

Свойство «двойственности» МЗ  $(A \rightarrow B|C)$  предопределяет развитие алгоритма нормализации по пути декомпозиции исходного отношения на два новых, где имеют место ФЗ  $A \rightarrow B$  и  $A \rightarrow C$  из МЗ, подлежащей устранению. На этом этапе возможны несколько исходов:

- (1) атрибуты  $A$ ,  $B$  и  $C$  являются абстракциями свойств одного и того же объекта учета;
- (2) атрибуты  $A$ ,  $B$  и  $C$  являются абстракциями свойств трех разнородных (обладающих различающимися наборами атрибутов) объектов учета;
- (3) атрибут  $A$  — абстракция свойства родительского отношения, а атрибуты  $B$  и  $C$  — абстракции свойств дочернего отношения;
- (4) атрибуты  $A$  и  $B$  являются абстракциями свойств одного и того же объекта учета, а атрибут  $C$  — абстракция свойства другого (не однородного первому) объекта учета.

Формально существует еще один вариант, когда атрибуты  $A$  и  $C$  являются абстракциями свойств одного и того же объекта учета, а атрибут  $B$  — абстракция свойства другого (не однородного первому) объекта учета. Но этот вариант тождествен предыдущему.

В первом случае, если выполнять декомпозицию исходного отношения на пару новых, возникнет необходимость их натурального соединения по общему атрибуту  $A$ . Так как натуральное соединение обеспечивает попарное сцепление всех возможных вариаций значений атрибутов при условии совпадения их значений, то возникнет эффект «ложных кортежей», приводящий к утрате целостности БД.

Во втором случае не будет достигнута цель приведения БД к 4НФ, так как ФЗ  $A \rightarrow B$  и  $A \rightarrow C$  на самом деле таковыми не являются, описывая ЗС между атрибутами (абстракциями значимых свойств) разнородных объектов учета внутри одного отношения. Фактически новые отношения не будут удовлетворять требованиям 4НФ.

В случае когда атрибут  $A$  — абстракция свойства родительского отношения, а атрибуты  $B$  и  $C$  — абстракции свойств дочернего отношения, опять возникнет необходимость соединения кортежей новых отношений, так как они содержат данные о разнородных объектах учета, что приведет к возникновению «ложных кортежей». Случай 4 схож с предыдущим случаем в части одного из новых отношений. Он также не может считаться удовлетворительным решением задачи приведения отношения к 4НФ.

Решение описанной проблемы, по мнению авторов, сводится к правильной декомпозиции атрибутов, составляющих нормализуемое отношение, а именно: никакие атрибуты нормализованного отношения не должны абстрагировать никакие свойства иных объектов учета, кроме абстрагируемого данным. Приведенное положение легко обосновывается самой природой реляционного отношения как следствия декартова (прямого) произведения множеств.

Исходя из абстрактного смысла связи (relationship), можно сформулировать семантическую интерпретацию нормальной формы, основанной на ЗС. Отношение приведено к 5НФ (проекционно-соединительной) тогда и только тогда, когда оно удовлетворяет требованиям 4НФ и детерминантами ЗС являются только потенциальные ключи и ничто иное. Выполнение данного требования возможно только в случае, когда все значимые свойства объекта учета абстрагируются атрибутами единственного отношения, а не распределены по более чем одному отношению.

## 5 Заключение

Значимая для обеспечения целостности БД процедура нормализации обладает рядом свойств, препятствующих ее сходимости, а следовательно, пригодности для успешного применения. Сформулированные в работе уточнения в виде семантических интерпретаций сложных и неочевидных правил декомпозиции отношений обеспечивают реализуемость теоретических правил в практической деятельности инженеров.

При формировании 1НФ необходимым условием ее достижения выступает результативное изыскание полного множества потенциальных ключей, их исследование на предмет тривиальности и минимальности (неприводимости слева). Последующие правила нормализации в большинстве своем оперируют качественными потенциальными ключами, а результативность и безошибочность их применения зависят от полноты и точности результата процедуры определения совокупности потенциальных ключей.

Снятие неопределенности относительно принадлежности атрибута или совокупности атрибутов к множеству потенциальных ключей достаточно интерпретировано исследованием изменчивости детерминанта и зависимой части в соответствии с представленной теоремой. Важным следствием данной теоремы является решающее правило разделения атрибутов отношения на ключевые (принадлежащие потенциальным ключам) и неключевые.

Научная новизна представленного исследования заключается в реализации оригинального семантического моделирования ФЗ, ЗС, потенциальных ключей, отношений реляционной БД, обеспечивающего повышение строгости выполнения правил приведения отношений к нормальным формам, вплоть до пятой, основанного на интерпретации понятия функции в терминах реляционных БД, формирования гипотез о влиянии значимых свойств ФЗ и ЗС на защищенность от аномалий обновления, избыточности и последствий декомпозиции с соединением с потерями.

Практическая ценность работы состоит в формулировке и доказательстве теоремы о ФЗ, обобщении и интерпретации правил декомпозиции отношений в процессе их нормализации и описании семантически детализированных приемов реализации данного процесса. Значимым, по мнению авторов, результатом представляется уточнение алгоритма нормализации в терминах инженерного применения, обеспечивающего формирование пригодных структур БД.

## **Литература**

1. *Date C. J., Darwen H.* Foundation for future database systems: The third manifesto. — 2nd ed. — Reading, MA, USA: Addison-Wesley, 2000. 547 p.
2. *Глухарев М. Л.* Применение формальных описателей для логико-алгебраического моделирования требований целостности информации в реляционных базах данных // Известия Петербургского университета путей сообщения, 2010. № 4(25). С. 78–88.
3. ГОСТ Р ИСО/МЭК ТО 10032-2007. Эталонная модель управления данными. — М.: Стандартинформ, 2009. 40 с.
4. *Козаловский М. Р.* Энциклопедия технологий баз данных. — М.: Финансы и статистика, 2002. 800 с.
5. Межгосударственный стандарт ГОСТ 33707-2016 (ISO/IEC 2382-2015). Информационные технологии. Словарь. — М.: Стандартинформ, 2017. 202 с.
6. *Цаленко М. Ш.* Моделирование семантики в базах данных. — М.: Наука, 1989. 288 с.
7. *Клименко И. В.* Метод формальной нормализации отношений реляционной модели базы данных // Вестник Ростовского государственного университета путей сообщения, 2012. № 2(46). С. 79–87.
8. *Рудометкина М. Н., Спицын В. Г., Эттель В. А.* Разработка реляционных средств декомпозиции предикатов // Известия Томского политехнического университета, 2013. Т. 322. № 5. С. 131–137.

9. *Arenas M., Libkin L.* An information-theoretic approach to normal forms for relational and XML data // J. ACM, 2003. Vol. 52. Iss. 2. P. 15–26. doi: 10.1145/773153.773155.
10. *Kumar K., Azad S. K.* Database normalization design pattern // 4th IEEE Uttar Pradesh Section Conference (International) on Electrical, Computer and Electronics. — Mathura, India: IEEE, 2017, P. 318–322. doi: 10.1109/UPCON.2017.8251067.
11. *Григорьев Ю. А., Устимов А. И.* Сравнение времени выполнения запроса к хранилищу данных в среде MapReduce/Hadoop и СУБД MySQL // Информатика и системы управления, 2016. № 3(49). С. 3–12. doi: 10.22250/isu.2016.49.3-12.
12. *Тараканов О. В., Пикалова Н. Б., Тараканова В. О.* Исследование «парадокса нормализации» в фактографических реляционных базах данных на платформе ORACLE // Научно-технический вестник информационных технологий, механики и оптики, 2018. Т. 18. № 4. С. 639–645. doi: 10.17586/2226-1494-2018-18-4-639-645.
13. *Кузнецов С. Д.* Основы баз данных. — 2-е изд. — М.: Интернет-университет информационных технологий; БИНОМ. Лаборатория знаний, 2007. 484 с.
14. *Pin-Shan Chen P.* The entity-relationship model — toward a unified view of data // ACM T. Database Syst., 1976. Vol. 1. No. 1. P. 9–36. doi: 10.1145/320434.320440.
15. *Мейер Д.* Теория реляционных баз данных / Пер. с англ. — М.: Мир, 1987. 608 с. (*Maier D.* The theory of relational databases. — Rockville, MD, USA: Computer Science Press, 1983. 637 p.)
16. *Bernstein P.* Synthesizing third normal form relations from functional dependencies // ACM T. Database Syst., 1976. Vol. 1. No. 4. P. 277–298. doi: 10.1145/320493.320489.
17. *Григорьев Ю. А.* Алгоритм синтеза частично оптимальной схемы реляционной базы данных // Наука и образование, 2012. № 1. Ст. 77–30569/294486.
18. *Асташина Е. А., Баранчиков А. И.* Реинжиниринг многозначных зависимостей и проблема избыточности данных // Перспективы развития информационных технологий, 2012. № 7. С. 9–13.
19. *Карпук А., Краснопрошин В. В.* Циклы в структурах функциональных зависимостей // Int. J. Open Information Technologies, 2017. Т. 5. № 7. С. 38–44.

Поступила в редакцию 13.09.22



## SEMANTIC INTERPRETATIONS OF HIGH NORMAL FORMS OF RELATIONS IN A RELATIONAL DATABASE

V. A. Ivanov<sup>1</sup>, M. Yu. Konyshov<sup>2</sup>, S. V. Smirnov<sup>1</sup>, O. V. Tarakanov<sup>1</sup>,  
V. O. Tarakanova<sup>3</sup>, and S. V. Usovik<sup>1</sup>

<sup>1</sup>MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation

<sup>2</sup>Financial University under the Government of the Russian Federation, 49 Leningradsky Prosp., Moscow 125993, Russian Federation

<sup>3</sup>Federal Guard Service of the Russian Federation, Moscow Kremlin, Moscow 109012, Russian Federation

**Abstract:** The results of semantic modeling of the processes of eliminating redundancy and protecting the relation of a relational database from update anomalies by improving the process of its normalization are presented. The semantic equivalents of the requirements of the normal forms of the relational database relations were established which increase the efficiency of the normalization algorithm. A theorem was formulated and proved that guarantees an unambiguous determination of the functional dependence of an attribute on a potential key. Refinements of the relational relation normalization algorithm were formulated. The ways of applying the proposed interpretations are given. Practically applicable procedures are shown that remove the contradictions of designing a rational database structure. The main findings are formulated in relation to factual databases. The proposed mechanisms are independent of the database management system used.

**Keywords:** relational database; normalization of relation; potential key; nonkey attribute; functional dependence; multivalued dependence

**DOI:** 10.14357/08696527230105

### References

1. Date, C. J., and H. Darwen. 2000. *Foundation for future database systems: The third manifesto*. 2nd ed. Reading, MA: Addison-Wesley. 547 p.
2. Glukharev, M. L. 2010. Primenenie formal'nykh opisateley dlya logiko-algebraicheskogo modelirovaniya trebovaniy tselostnosti informatsii v relyatsionnykh bazakh dannykh [Application of formal specifiers for logic-algebraic modeling of information integrity demands in relational databases]. *Izvestiya Peterburgskogo universiteta putey soobshcheniya* [Proceedings of Petersburg Transport University] 4(25):78–88.
3. GOST R ISO/MEK TO 10032-2007. 2009. Etalonnaya model' upravleniya dannymi [Data management reference model]. Moscow: Standartinform Pubs. 40 p.
4. Kogalovskiy, M. R. 2002. *Entsiklopediya tekhnologii baz dannykh* [Database technology encyclopedia]. Moscow: Finance and Statistics Pubs. 800 p.
5. GOST 33707-2016 (ISO/IEC 2382-2015). 2017. Mezhgosudarstvennyy standart. Informatsionnye tekhnologii. Slovar' [Interstate standard. Information technologies. Vocabulary]. Moscow: Standartinform Pubs. 202 p.



6. Tsalenko, M. Sh. 1989. *Modelirovanie semantiki v bazakh dannykh* [Modeling semantics in databases]. Moscow: Nauka. 288 p.
7. Klimenko, I. V. 2012. Metod formal'noy normalizatsii otnosheniy relyatsionnoy modeli bazy dannykh [Method for formalized normalization the relations of relational model]. *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov State University of Transport] 2(46):79–87.
8. Rudometkina, M. N., V. G. Spitsyn, and V. A. Ettl. 2013. Razrabotka relyatsionnykh sredstv dekompozitsii predikativ [Development of tools for relational decomposition of predicates]. *Bulletin of the Tomsk Polytechnic University* 322(5):131–137.
9. Arenas, M., and L. Libkin. 2003. An information-theoretic approach to normal forms for relational and XML data. *J. ACM* 52(2):15–26. doi: 10.1145/773153.773155.
10. Kumar, K., and S. K. Azad. 2017. Database normalization design pattern. *4th IEEE Uttar Pradesh Section Conference (International) on Electrical, Computer and Electronics*. Mathura, India: IEEE. 318–322. doi: 10.1109/UPCON.2017.8251067.
11. Grigorev, Yu. A., and A. I. Ustimov. 2016. Sravnenie vremeni vypolneniya zaprosa k khranilishchu dannykh v srede MapReduce/Hadoop i SUBD MySQL [Comparison of query execution time to data warehouse in MapReduce/Hadoop and RDBMS MySQL]. *Informatika i sistemy upravleniya* [Information Science and Control Systems] 3(49):3–12. doi: 10.22250/isu.2016.49.3-12.
12. Tarakanov, O. V., N. B. Pikalova, and V. O. Tarakanova. 2018. Issledovanie “paradoksa normalizatsii” v faktograficheskikh relyatsionnykh bazakh dannykh na platforme ORACLE [“Normalization paradox” research in factual relational databases on ORACLE platform]. *Nauchno-tekhnicheskiiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical J. of Information Technologies, Mechanics and Optics] 18(4):639–645. doi: 10.17586/2226-1494-2018-18-4-639-645.
13. Kuznetsov, S. D. 2007. *Osnovy baz dannykh* [Database fundamentals]. Moscow: Internet-University of Information Technology; BINOM. Knowledge laboratory. 484 p.
14. Pin-Shan Chen, P. 1976. The entity-relationship model — toward a unified view of data. *ACM T. Database Syst.* 1(1):9–36. doi: 10.1145/320434.320440.
15. Maier, D. 1983. *The theory of relational databases*. Rockville, MD: Computer Science Press. 637 p.
16. Bernstein, P. 1976. Synthesizing third normal form relations from functional dependencies. *ACM T. Database Syst.* 1(4):277–298. doi: 10.1145/320493.320489.
17. Grigor'ev, Yu. A. 2012. Algoritm sinteza chastichno optimal'noy skhemy relyatsionnoy bazy dannykh [Algorithm of synthesis of suboptimal scheme of relational database]. *Nauka i obrazovanie* [Science and Education] 1:77–30569/294486.
18. Astashina, E. A., and A. I. Baranchikov. 2012. Reinzhiniring mnogoznachnykh zavisimostey i problema izbytochnosti dannykh [Reengineering of multivalued dependencies and the problem of data redundancy]. *Perspektivy razvitiya informatsionnykh tekhnologiy* [Prospects for the Development of Information Technologies] 7:9–13.
19. Karpuk, A. A., and V. V. Krasnoproshin. 2017. Tsikly v strukturakh funktsional'nykh zavisimostey [Cycles in structures of functional dependencies]. *Int. J. Open Information Technologies* 5(7):38–44.

Received September 13, 2022

## **Contributors**

**Ivanov Vladimir A.** (b. 1957) — Doctor of Military Science, professor, Corresponding Member of the Academy of Cryptography of the Russian Federation; professor, MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; iva.mac@mail.ru

**Konyshov Mikhail Yu.** (b. 1975) — Doctor of Science in technology, professor, Financial University under the Government of the Russian Federation, 49 Leningradsky Prosp., Moscow 125993, Russian Federation; my-konyshov@fa.ru

**Smirnov Sergei V.** (b. 1976) — Candidate of Science (PhD) in technology, associate professor, MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; smirnov\_s@mirea.ru

**Tarakanov Oleg V.** (b. 1966) — Candidate of Science (PhD) in technology, associate professor, MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; ole\_g66@list.ru

**Tarakanova Valeria O.** (b. 1997) — employee, Federal Guard Service of the Russian Federation, Moscow Kremlin, Moscow 109012, Russian Federation; talero@list.ru

**Usovik Sergei V.** (b. 1982) — Candidate of Science (PhD) in technology, assistant professor, MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; usovik@mirea.ru

## ЗАДАЧА КЛАССИФИКАЦИИ В УСЛОВИЯХ ИСКАЖЕННЫХ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ

*А. А. Грушо<sup>1</sup>, Н. А. Грушо<sup>2</sup>, М. И. Забейхайло<sup>3</sup>, А. А. Зацаринный<sup>4</sup>,  
Е. Е. Тимонина<sup>5</sup>, С. Я. Шоргин<sup>6</sup>*

**Аннотация:** Рассматривается модель классификации потока объектов на предмет наличия или отсутствия в каждом объекте  $O$  некоторого свойства  $A$ . Предполагается, что существуют  $M$  взаимно однозначных преобразований объектов, поступающих на классификацию, и в потоке встречаются объекты, полученные из  $O$  по одному из этих преобразований. Для объекта  $O$  известно, что в нем содержится свойство  $A$ , которое послужило причиной появления известных объектов  $B_1, B_2, \dots, B_k$  в информационных пространствах  $I_1, I_2, \dots, I_k$ . Это означает, что выявить свойство  $A$  можно только наблюдая следствия  $B_1, B_2, \dots, B_k$ . Задача заключается в том, что для каждого приходящего в потоке объекта необходимо определить присутствие или отсутствие в нем преобразованной причины  $A$ . Построены алгоритмы проверки такой возможности в случаях, когда есть описание характеристик  $A$  и когда такое описание отсутствует.

**Ключевые слова:** задача конечной классификации; причинно-следственные связи; машинное обучение в условиях искажений

**DOI:** 10.14357/08696527230106

### 1 Введение

Использование причинно-следственных связей для распознавания и 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 объектов, появление которых не имеет причин. Если причина искажена, то следствие либо не возникает, либо обусловлено другими причинами.

---

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru

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

<sup>3</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabekhailo@yandex.ru

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

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

<sup>6</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

Применение причинно-следственных связей в задачах классификации исследовалось во многих научных работах. Первые прикладные исследования и теоретические обоснования использования причинно-следственных связей в данных задачах относятся к медицинской тематике. В работах [1–3] построены модели каузального вывода в медицинской диагностике. В [4] построена модель использования причинно-следственных связей в задачах классификации изображений с применением байесовских и нейронных сетей. В этой работе показано, что качество классификации зависит от качества реальных входных данных классификации, детерминированных сдвигов изображения по горизонтали и вертикали и случайных помех. При этом сдвиги служат причинами фактического снижения качества распознавания объектов на изображениях, а именно: сдвиги становятся причинами нарушения устойчивости методов распознавания в отличие от шума, имеющего иное распределение вероятностей. Используя эти различия распределений, авторы статьи построили методы и оценки сдвигов в изображениях. Это позволило учитывать при классификации оценки возможных сдвигов, что резко повысило устойчивость результатов классификации.

В данной работе детерминированность появления следствия из причины используется следующим образом. При решении проблемы бинарной классификации в потоке однотипных объектов определяется наличие или отсутствие причины по наблюдениям за информационными пространствами. В этих пространствах проявляются следствия появления или отсутствия искомой причины в классифицируемых объектах. Однотипность объектов предполагает, что известны виды изменений характеристик поступающих объектов относительно некоторого канонического объекта.

Однако изменения характеристик причины порождают нарушение причинно-следственных связей, т. е. следствия могут появляться только из-за шума. Тогда задача определения причины в измененном объекте должна решаться с использованием знаний о возможных преобразованиях, а именно: необходимо опробовать изменения таким образом, чтобы добиться совпадения с каноническим объектом и по этому совпадению принимать решение о наличии причины по появившимся детерминированным следствиям. Решение задачи поиска причины в поступающих объектах рассмотренным методом потребовало построения моделей, в которых можно получать строгие результаты.

## **2 Проблема классификации при изменениях причин**

Рассмотрим задачу бинарной классификации. Пусть необходимо неоднократно классифицировать типовые объекты  $O$  на предмет наличия или отсутствия в них некоторого свойства  $A$ . Будем считать, что существует множество  $U = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$  характеристик и любой объект  $O$  описывается подмножеством множества  $U$ . Свойство  $A$  определяется некоторым подмножеством объекта  $O$ , но состав его характеристик может быть неизвестен.

Как было принято в работе [5], следствия  $B_1, B_2, \dots, B_k$  порождаются свойством  $A$  как причиной ( $A$  — причина появления следствий  $B_1, B_2, \dots, B_k$ ) и являются объектами в соответствующих информационных пространствах  $I_1, I_2, \dots, I_k$ . В данной работе классификация рассматривается как наличие или отсутствие свойства  $A$  по появлению одного или нескольких его следствий  $B_1, B_2, \dots, B_k$ . Любой алгоритм классификации принимает решение о классе объекта, наблюдая идентифицируемые следствия  $B_1, B_2, \dots, B_k$  свойства  $A$  в  $O$ , т. е. опосредованно, путем их появления в информационных пространствах  $I_1, I_2, \dots, I_k$  в случае присутствия  $A$  в объекте  $O$ .

В работе [5] следствия  $B_1, B_2, \dots, B_k$  называются дополнительной информацией, которая может подтвердить наличие свойства  $A$  в объекте  $O$ . Причина  $A$  (свойство  $A$ ) определяется как минимальное множество характеристик, при которых в пространствах  $I_1, I_2, \dots, I_k$  детерминированно появляются следствия  $B = (B_1, B_2, \dots, B_k)$ . Это значит, что существует детерминированный алгоритм  $h$ , который преобразует  $A$  в  $B$ . При этом на следствия возможно воздействие независимых шумов, которые могут искажать характеристики следствий. Поскольку истинные значения следствий можно считать известными, то определение наличия следствия или его отсутствия можно решать статистическими средствами [5].

Поток однотипных объектов  $O$ , подлежащих классификации, и классифицируемые в них свойства (причины)  $A$  могут изменяться. В поиске конкретного свойства в поступающем объекте возможность его изменения связана с тем, что подобные условия классификации возникают неоднократно и в каждом случае некоторые характеристики объекта могут меняться на другие, но новый поиск системы следствий является трудоемкой задачей. Из определения причины следует, что при изменении (отсутствии) хотя бы одной характеристики причины она перестает быть причиной для следствия [6]. Поэтому при преобразовании объекта  $O$  таким образом, что меняется состав характеристик причины  $A$ , исчезает возможность использования следствий  $A$  для обоснования наличия  $A$  в  $O$ .

Рассмотрим пример, в котором в изображениях  $D$  ищется начерченный прямоугольник, называемый причиной. Если прямоугольник присутствует в  $D$ , то целесообразно искать сначала отрезки прямых линий (рис. 1 и 2). Эти отрезки являются следствиями наличия прямоугольника в изображении, и существуют эффективные алгоритмы поиска отрезков прямых линий. При появлении нового



**Рис. 1** Пример распознавания прямоугольника (вариант 1)



**Рис. 2** Пример распознавания прямоугольника (вариант 2)

изображения для такого же прямоугольника может меняться его сдвиг или поворот.

Предположим сначала, что форма прямоугольника известна. Однако его поиск на изображении — сложная задача, которую целесообразно решать, выявляя сначала сегменты (стороны) прямоугольника. При повороте прямоугольника сегменты (см. вариант 2 на рис. 2) меняются, т. е. причина их появления не может быть той же, что и в варианте 1 на рис. 1.

### 3 Причины и следствия в измененных объектах

Пусть причина  $A$  однозначно описывается в полученном объекте  $O$  конечным множеством характеристик  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ . В другом подобном объекте  $O^1$  имеем другой вектор характеристик  $\alpha^1 = (\alpha_1^1, \alpha_2^1, \dots, \alpha_n^1)$ , который получается с помощью преобразования  $T$  исходного вектора  $\alpha$ . Пусть  $T(A) = A_1$ , тогда наличие  $A_1$  в объекте  $O^1$  необходимо проверять исходя из того, что  $A_1$  порождает следствия  $B_1^1, B_2^1, \dots, B_s^1$ , если они существуют.

Предположим, что существуют  $M$  возможных преобразований  $T$ , которые порождают новые условия классификации. Будем считать, что все преобразования  $T$  являются изоморфизмами  $A$ . Тогда существует детерминированный алгоритм  $T^{-1}(A_1) = A$ . Из детерминированности алгоритмов  $T^{-1}$  и  $h$  следует, что их суперпозиция также будет детерминированным алгоритмом, т. е. преобразование  $h(T^{-1}(A_1)) = B$ , причем исключение хотя бы одной характеристики из  $A_1$  исключает какие-то характеристики из  $A$ , что делает это равенство неверным. Таким образом,  $A_1$  служит причиной множества следствий  $B$ , что не нарушает свойство единственности причин, так как  $A$  и  $A_1$  находятся в разных пространствах. Проведенные рассуждения суммируются в следующем утверждении.

**Утверждение.** *Если все преобразования  $A$  являются изоморфизмами и алгоритмы  $T^{-1}$  и  $h$  построены, то для всех  $M$  преобразований  $T$  множество  $B$  является следствием причин  $T(A)$ .*

Рассмотренные выше случаи предполагали знание характеристик причины  $A$ . Вместе с тем возможно, что характеристики  $A$  неизвестны, но известно, что характеристики причины  $A$  могут быть покрыты известным объектом  $O$ . В этом случае также можно проводить классификацию на предмет наличия в поступившем объекте причины  $A$ , т. е.  $A$  детерминированно порождает множество следствий  $B$ , и если эти следствия известны, то их появление в соответствующих информационных пространствах  $I_1, I_2, \dots, I_k$  позволит сделать вывод о наличии  $A$  в  $O$ .

Как в предыдущем случае, пусть  $T_1, T_2, \dots, T_M$  — преобразования  $O$ , которые могут появиться в потоке классифицируемых данных. Предположим, что эти преобразования являются изоморфизмами. Тогда алгоритмы обращения  $T_1^{-1}, T_2^{-1}, \dots, T_M^{-1}$  детерминированно отображают очередной поступивший

объект классификации в исходный объект  $O$ , а детерминированные следствия наличия причины  $A$  порождают следствия  $B$  модифицированных объектов  $O$ .

Вернемся к рассмотрению примера. В нем использована дополнительная информация, что следствия получены сегментацией исходного прямоугольника. Добавим дополнительное условие, что преобразование  $T(A) = A_1$  сохраняет сегментацию, оставаясь изоморфизмом. Сохранение сегментации означает, что каждый сегмент  $A$  переходит в изоморфный сегмент в  $A_1$ . Отображение  $h$  может быть взято в форме тождественных преобразований сегментов  $A$  на себя. Тогда отображение  $h^*$  тождественно выделяет сегменты  $A_1$ , которые изоморфны сегментам  $h(A)$  и являются следствиями причины  $A_1$ . В самом деле,  $A_1$  детерминированно порождает сегментацию  $h^*(A_1)$ , которая при обратном отображении порождает следствия  $B$ , как это делалось раньше.

Предположим, что  $h^*(A_1)$  не является следствием причины  $A_1$ , т.е. при удалении хотя бы одной характеристики  $A_1$  в силу изоморфизма удаляется изоморфная характеристика в  $A$ . Тогда  $A$  перестает быть причиной следствий  $h(A)$ , что противоречит условию. Эти рассуждения показывают, что следствия  $h^*(A_1)$  детерминированно порождены характеристиками  $A_1$  и эти характеристики образуют минимальный набор. Отсюда следует, что между следствиями существует взаимно однозначное отображение  $\tau$  сегментов  $h(A)$  в сегменты  $h^*(A_1)$ .

Если алгоритм  $T$  и алгоритмы  $h$  и  $h^*$  не построены, то приведенные рассуждения показывают, как этих построений можно избежать. Во-первых, выше показано, что  $h^*(A_1)$  являются следствиями причины  $A_1$ , которая изоморфна причине  $A$ , т.е. наличие прямоугольника  $A_1$  следует из  $h^*(A_1)$  так же, как наличие прямоугольника  $A$  следует из  $h(A)$ . Во-вторых, можно построить само преобразование  $A$  в  $A_1$ . Для этого нужно эмпирически построить преобразование  $\tau$  между сегментами  $h(A)$  и сегментами  $h^*(A_1)$ . Для рассмотренного примера это значит, что надо определить углы наклона между следствиями на рис. 1 и 2. Если такие углы удастся построить, то из следствий будет вытекать, что в обоих изображениях присутствует искомый прямоугольник и какой поворот надо осуществить, чтобы из одного получить другой.

Таким образом, при сегментации изменения положения объекта классификации порождают изменения расположения сегментов, что позволяет осуществлять классификацию по сегментам при изменениях их положения.

## 4 Алгоритмы классификации по следствиям

Предположим, что число  $M$  преобразований объекта классификации велико и заранее не известно, какому из преобразований исходного объекта  $O$  соответствует очередной объект  $O^1$ . Понятно, что существует только одно истинное преобразование, которое может детерминированно породить следствия из классифицируемой причины, если она присутствует в  $O^1$ . Для проверки наличия следствий причины  $A_1$  необходимо сделать обратное преобразование  $T^{-1}$  и для



полученного результата вычислить алгоритм  $h$ , чтобы в информационных пространствах  $I_1, I_2, \dots, I_k$  искать следствия  $B_1, B_2, \dots, B_k$ . Если преобразование  $T^{-1}$  выбрано неправильно, то некоторые следствия могут появиться только за счет шума.

При больших  $M$  и  $k$  задача сводится к поиску вкраплений в большие данные. Эта проблема широко исследовалась в научной литературе [6, 7]. В [7–11] построен эффективный метод решения некоторых подобных задач. Если преобразование  $A$  в  $A_1$  описывается треугольной матрицей, то идея метода состоит в том, чтобы заменить переборный процесс на поиск с помощью последовательности упрощенных моделей.

## 5 Заключение

В работе построены методы бинарной классификации потока однотипных событий на основе причинно-следственных свойств в данных. Основная сложность решенной задачи заключалась в том, что однотипность событий не означает их одинаковость. Это значит, что даже при небольших изменениях в классифицируемых объектах может меняться причина следствий, на которых строится вывод о наличии этой причины. Поэтому потребовалось доказательство того, что объекты одного типа также можно использовать для причинно-следственной классификации.

Рассмотрены несколько типов условий классификаций на основе причинно-следственных связей.

Первый тип предполагает знание описания причины в эталонном объекте и вид преобразования этой причины в однотипных объектах.

Второй тип не предполагает знание причины, но известно, что причина во всех объектах покрывается известным объектом и известны преобразования объектов, содержащих причины.

Третий тип связан с распознаванием объекта в изображении. Для сокращения вычислений и устойчивости решения объект поиска ассоциируется с причиной, а следствия причины представляют собой сегменты причины. В этом случае доказано, что классификацию однотипных объектов можно проводить, отождествляя сегменты эталонного объекта с сегментами нового объекта.

## Литература

1. *Pearl J.* Causal inference // Workshop on Causality Proceedings: Objectives and Assessment at NIPS / Eds. I. Guyon, D. Janzing, B. Scholkopf. — Proceedings of machine learning research ser. — Whistler, Canada, 2010. Vol. 6. P. 39–58.
2. *Richens J. G., Lee C. M., Johri S.* Improving the accuracy of medical diagnosis with causal machine learning // Nat. Commun., 2020. Vol. 11. Art. 3923. 9 p. doi: 10.1038/s41467-020-17419-7.



3. *Забейайло М. И., Грушо А. А., Грушо Н. А., Тимонина Е. Е.* Поддержка решения задач диагностического типа // Системы и средства информатики, 2021. Т. 31. № 1. С. 69–81.
4. *Zhang C., Zhang K., Li Y.* A causal view on robustness of neural networks // 34th Conference on Neural Information Processing Systems. — Cornell University, 2021. arXiv: 2005.01095v3 [cs.LG].
5. *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.
6. *Грушо А. А., Грушо Н. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я.* Поиск аномалий в больших данных // Системы и средства информатики, 2022. Т. 32. № 1. С. 160–167. doi: 10.14357/08696527220115.
7. *Vaughan G.* Efficient big data model selection with applications to fraud detection // Int. J. Forecasting, 2018. Vol. 36. Iss. 3. P. 1116–1127. doi: 10.1016/j.ijforecast.2018.03.002.
8. *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.
9. *Смирнов Д. В., Грушо А. А., Забейайло М. И., Тимонина Е. Е.* Система сбора и анализа информации из различных источников в условиях Big Data // Int. J. Open Information Technologies, 2021. Vol. 9. No. 4. P. 64–71.
10. *Смирнов Д. В.* Методика проблемно-ориентированного анализа Big Data в режиме ограниченного времени // Int. J. Open Information Technologies, 2021. Vol. 9. No. 9. P. 88–94.
11. *Грушо А. А., Грушо Н. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я.* Статистика и кластеры в поисках аномальных вкраплений в условиях больших данных // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 79–86.

*Поступила в редакцию 13.02.23*

---

## CLASSIFICATION PROBLEM IN CONDITIONS OF DISTORTED CAUSE-AND-EFFECT RELATIONSHIPS

*A. A. Grusho, N. A. Grusho, M. I. Zabezhailo, A. A. Zatsarinny, E. E. Timonina,  
and S. Ya. Shorgin*

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

**Abstract:** The paper considers the model for classifying an object stream for the presence or absence of a certain property  $A$  in each object  $O$ . It is assumed that there are  $M$  bijective transformations of objects coming for classification and in the stream, there are objects obtained from  $O$  according to one of these transformations. For each object  $O$ , it is known that it contains property  $A$

which causes the known objects  $B_1, B_2, \dots, B_k$  to appear in information spaces  $I_1, I_2, \dots, I_k$ . This means that property  $A$  can only be detected by observing the consequences  $B_1, B_2, \dots, B_k$ . The problem is that for each object in the flow, it is necessary to determine the presence or absence of the converted reason  $A$  in it. The algorithms for checking such possibility are built in cases where there is a description of characteristics  $A$  and when such a description is absent.

**Keywords:** finite classification task; cause-and-effect relationships; machine learning in distortions conditions

**DOI:** 10.14357/08696527230106

## References

1. Pearl, J. 2010. Causal inference. *Workshop on Causality Proceedings: Objectives and Assessment at NIPS*. Eds. I. Guyon, D. Janzing, and B. Scholkopf. Proceedings of machine learning research ser. Whistler, Canada. 6:39–58.
2. Richens, J. G., C. M. Lee, and S. Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11(1):3923. 9 p. doi: 10.1038/s41467-020-17419-7.
3. Zabezhalo, M. I., A. A. Grusho, N. A. Grusho, and E. E. Timonina. 2021. Podderzhka resheniya zadach diagnosticheskogo tipa [Support for solving diagnostic type problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):69–81.
4. Zhang, C., K. Zhang, and Y. Li. 2021. A causal view on robustness of neural networks. *34th Conference on Neural Information Processing Systems*. Cornell University. Available at: <https://arxiv.org/pdf/2005.01095.pdf> (accessed February 5, 2023).
5. Grusho, A. A., N. A. Grusho, M. I. Zabezhalo, 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.
6. Grusho, A. A., N. A. Grusho, M. I. Zabezhalo, D. V. Smirnov, E. E. Timonina, and S. Ya. Shorgin. 2022. Poisk anomalii v bol'shikh dannykh [Search of anomalies in big data]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):160–167. doi: 10.14357/08696527220115.
7. Vaughan, G. 2018. Efficient big data model selection with applications to fraud detection. *Int. J. Forecasting* 36(3):1116–1127. doi: 10.1016/j.ijforecast.2018.03.002.
8. Grusho, A. A., N. A. Grusho, M. I. Zabezhalo, and E. E. Timonina. 2020. Lokalizatsiya iskhodnoy prichiny anomalii [Root cause anomaly localization]. *Problemy informatsionnoy bezopasnosti. Komp'yuternyye sistemy* [Problems of Information Security. Computer Systems] 4:9–16.
9. Smirnov, D. V., A. A. Grusho, M. I. Zabezhalo, and E. E. Timonina. 2021. Sistema sbora i analiza informatsii iz razlichnykh istochnikov v usloviyakh Big Data [System for collecting and analyzing information from various sources in Big Data conditions]. *Int. J. Open Information Technologies* 9(4):64–71.
10. Smirnov, D. V. 2021. Metodika problemno-orientirovannogo analiza Big Data v rezhime ogranichenogo vremeni [Methodology of problem-oriented Big Data analysis in limited time mode]. *Int. J. Open Information Technologies* 9(9):88–94.

11. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, E. E. Timonina, and S. Ya. Shorgin. 2021. Statistika i klasteriy v poiskakh anomal'nykh vkrapleniy v usloviyakh bol'shikh dannyykh [Statistics and clusters for detection of anomalous insertions in Big Data environment]. *Informatika i ee Primeneniya* — *Inform. Appl.* 15(4):79–86.

*Received February 13, 2023*

## Contributors

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, 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, Institute of Informatics Problems, 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, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 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 119333, Russian Federation; AZatsarinny@ipiran.ru

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

**Shorgin Sergey Ya.** (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; sshorgin@ipiran.ru

## МЕТОДЫ КЛАССИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ СДО В МОДЕЛИ ПОСТРОЕНИЯ ИХ ИНДИВИДУАЛЬНОЙ ТРАЕКТОРИИ ОБУЧЕНИЯ\*

*Я. Г. Мартюшова<sup>1</sup>, Т. А. Минеева<sup>2</sup>, А. В. Наумов<sup>3</sup>*

**Аннотация:** Исследуется проблема адаптации системы дистанционного обучения (СДО) под контингент пользователей путем формирования индивидуальных траекторий обучения на основе результатов их промежуточных тестирований. Важнейшая составляющая предлагаемой модели — классификация пользователей на основе теста по различным категориям успеваемости. Проводится сравнительный анализ результатов применения для этой цели различных классификаторов, таких как байесовский классификатор, логистическая регрессия, метод *k*-ближайших соседей (KNN), дерево решений, метод случайного леса, градиентный бустинг, а также классификатора, использующего композицию нескольких различных алгоритмов (метод бутстреп-агрегирования, или бэггинг) из названных выше для предсказания категории пользователя по принципу простого большинства. Приводятся результаты численного эксперимента с использованием данных функционирования системы дистанционного обучения МАИ CLASS.NET.

**Ключевые слова:** система дистанционного обучения; методы машинного обучения; адаптивные системы; индивидуальная траектория обучения

**DOI:** 10.14357/08696527230107

### 1 Введение

Адаптация современных СДО под условия эксплуатации, в частности контингент пользователей, работающий в системе, становится важным направлением исследований [1–3]. Построение индивидуальных траекторий пользователей СДО — распространенный инструмент адаптации [4–6]. Одним из направлений исследований стало формирование индивидуальной траектории на основе результатов промежуточных тестирований и информации о работе пользователя путем его классификации в определенные моменты времени в различные категории по успеваемости. В зависимости от категории пользователя ему, например, могут предлагаться для дальнейшего изучения курса задания соответствующего уровня

---

\* Исследование выполнено за счет гранта Российского научного фонда № 22-28-00588 (<https://rscf.ru/project/22-28-00588>).

<sup>1</sup>Московский авиационный институт, [ma1554@mail.ru](mailto:ma1554@mail.ru)

<sup>2</sup>Московский авиационный институт, [mineeva.t2015@gmail.com](mailto:mineeva.t2015@gmail.com)

<sup>3</sup>Московский авиационный институт, [naumovav@mail.ru](mailto:naumovav@mail.ru)

сложности, определяемого экспертно, или на основе соответствующих математических моделей, например модели Раша [7]. Классификация пользователей СДО может осуществляться различными методами и на основе различных показателей работы. В качестве такого показателя в [4] при очной форме проведения занятий с организацией самостоятельной работы студентов и выполнения домашних работ в СДО используется текущий рейтинг пользователя, представляющий собой взвешенную свертку оценок за прошедшие очные контрольные работы, числа успешно выполненных разделов электронного учебника (ЭУ), числа пропущенных очных семинарских занятий и оценки, полученной студентом за коллоквиум. Однако вместо свертки как скалярного показателя можно использовать для классификации вектор из слагаемых свертки, как это сделано в [5, 6]. Для классификации пользователей используют различные методы машинного обучения, например байесовский классификатор [4], карты Кохонена [5, 6, 8], или другие методы, построенные на основе нейронных сетей [9–13].

Цель данной работы — сравнительный анализ различных методов классификации, используемых при построении индивидуальной траектории пользователя и конструирование нового объединенного классификатора, работающего по принципу большинства на основе анализа ансамбля классификаторов, построенных в том числе с использованием нейронных сетей. Использование объединенного классификатора обеспечивает большую стабильность корректной классификации, что подтверждается на основе обработки значительного массива данных о работе пользователей СДО МАИ CLASS.NET [14, 15], активно эксплуатируемой в Московском авиационном институте на протяжении последних двух десятков лет.

## **2 Описание математической модели построения индивидуальной траектории пользователя системы дистанционного обучения**

В данной работе продолжается исследование рассмотренной в [4] модели, учитывающей смешанную форму обучения в виде очных семинарских занятий в семестре с проведением нескольких объективных очных рубежных контролей в течение курса и организацией самостоятельной работы студентов с выполнением ими домашних заданий в СДО.

Предлагаемый пользователю курс дистанционного обучения разбит на несколько разделов, которые он должен изучить для освоения данной дисциплины. После каждого раздела пользователю нужно решить тест определенного уровня сложности. Сложность заданий теста определяется экспертом или с помощью специальных алгоритмов, основанных на модели Раша [7]. Для формирования теста из существующих заданий необходимо решить задачу математического программирования, описанную в [16], параметром которой служит суммарный уровень сложности теста. Необходимо определить, какого уровня сложности тест выдавать пользователю при проведении очных контрольных мероприятий,

чтобы он соответствовал уровню его подготовки и формировал индивидуальную траекторию обучения.

Решение о том, какой должна быть сложность выдаваемого теста, принимается экспертом-педагогом с использованием соответствующих психолого-педагогических методов [17] на основе того, к какой категории принадлежит пользователь (например, «отличник», «хорошист» и т. д.). Таким образом, на каждом этапе обучения предлагается решать задачу классификации пользователя в одну из категорий. Определяются категории пользователя дискретно в процессе обучения по окончании каждого раздела курса дистанционного обучения, который, как правило, соответствует очередному семинарскому занятию и очередному разделу в ЭУ. Обозначим моменты времени, когда проводится прогнозирование объективной категории пользователя, через  $t = 1, \dots, T$ .

В начальный момент времени  $t = 0$  объективной категории пользователя присваивается некоторая априорная оценка, определяемая на основе предыдущей траектории обучения. Пусть  $t = \{t_1, t_2, t_3\}$  — моменты времени, когда проводятся очные контрольные мероприятия, на основании которых проводится объективная классификация пользователей СДО с последующей адаптацией их индивидуальных траекторий обучения.

В моменты времени  $t_1, t_2$  и  $t_3$  объективная категория пользователя определяется на основе результата очной контрольной работы с помощью традиционной оценки  $k_{t_i} \in Y, i = 1, 2, 3$ , т. е.  $y_{t_i} = k_{t_i}, i = 1, 2, 3$ , где  $k_{t_i} \in Y = \{0, 2, 3, 4, 5\}$ . Случай  $k_{t_i} = 0$  означает, что пользователь не писал контрольную работу в момент времени  $t_i$ . В другие моменты времени объективная категория пользователя не определена.

В работе [4] в остальные моменты времени на основе текущего рейтинга пользователя определялась его текущая категория, а затем решалась задача классификации с целью построения прогноза будущей объективной категории пользователя. Для классификации использовался байесовский метод на основе анализа скалярного показателя работы пользователя — его текущего рейтинга, определяемого в момент времени  $t = 1, \dots, T$  следующим образом:

$$B_t = \omega_k \frac{\sum_{t_i \leq t} k_{t_i}}{5} + \omega_E \frac{\sum_{\tau=1}^t E_{\tau}^+}{\sum_{\tau=1}^t E_{\tau}} - \omega_{\delta} \left( 1 - \frac{\delta_t}{\Delta_t} \right),$$

где  $E_{\tau}$  — суммарная сложность теста, выполненного в ЭУ в момент  $\tau$ ;  $E_{\tau}^+$  — суммарная сложность правильно решенных заданий теста ЭУ, выполненного в момент  $\tau$ ;  $\Delta_t$  — общее число очных занятий, прошедших к моменту времени  $t$ ;  $\delta_t$  — число посещений пользователем очных занятий, прошедших к моменту времени  $t$ .

Величины  $\omega_k, \omega_E$  и  $\omega_{\delta}$  являются весовыми коэффициентами, которые выбраны так, чтобы максимальное значение текущего рейтинга пользователя в момент времени  $T$  было бы равно 100. Эти величины отражают важность каждого

слагаемого в формуле рейтинга. Например, в [17] выбраны следующие значения:  $\omega_k = \omega_E = 25$  и  $\omega_\delta = 15$ .

### 3 Построение обобщенного классификатора

Как уже было сказано, в [4] классификация пользователей проводилась по одному признаку, текущему рейтингу. В данной работе предлагается прогнозировать объективную категорию пользователя в момент времени  $t$  по нескольким признакам: нормированным суммарным сложностям правильно решенных заданий тестов, выполненных в моменты времени  $\tau \leq t$ , оценкам  $k_{t_i}$ ,  $i = 1, 2, 3$ , за очные контрольные работы, проводившиеся в моменты времени  $t_i < t$  (будем предполагать, что очные контрольные работы при  $T = 17$  проводятся на 5-м, 10-м и 15-м занятиях, т. е.  $t_i \in \{5, 10, 15\}$ ), долям пропущенных пользователем очных занятий, прошедших в период до момента времени  $t$ . Например, если решение задачи классификации происходит после 7-го раздела учебника, то классификация будет проводиться по следующим признакам:

- нормированные сложности тестов ЭУ в моменты  $\tau = 1, \dots, 7$ :

$$\frac{E_1^+}{E_1}, \frac{E_2^+}{E_2}, \frac{E_3^+}{E_3}, \frac{E_4^+}{E_4}, \frac{E_5^+}{E_5}, \frac{E_6^+}{E_6}, \frac{E_7^+}{E_7},$$

где  $E_\tau$  — суммарная сложность теста ЭУ, выполненного в момент  $\tau = 1, \dots, 7$ ;  $E_\tau^+$  — суммарная сложность правильно решенных заданий теста ЭУ, выполненного в момент  $\tau = 1, \dots, 7$ ;

- доли пропущенных занятий:

$$\frac{\theta_1}{\Delta_1}, \frac{\theta_2}{\Delta_2}, \frac{\theta_3}{\Delta_3}, \frac{\theta_4}{\Delta_4}, \frac{\theta_5}{\Delta_5}, \frac{\theta_6}{\Delta_6}, \frac{\theta_7}{\Delta_7},$$

где  $\theta_\tau$  — число пропущенных пользователем очных занятий, прошедших к моменту времени  $\tau = 1, \dots, 7$ ;  $\Delta_\tau$  — общее число очных занятий, прошедших к моменту времени  $\tau = 1, \dots, 7$ ;

- оценка за первую контрольную работу, которая проводилась после 5-го раздела электронного учебника:  $k_{t_1}$ .

В ходе данной работы для прогнозирования объективной категории применялись следующие классификаторы: логистическая регрессия, метод ближайших соседей (KNN), дерево решений, случайный лес, градиентный бустинг.

В качестве исходных данных были взяты результаты работы пользователей ЭУ по курсу «Теория вероятностей и математическая статистика» в СДО CLASS.NET, ранее использованные в [4]. Число моментов времени, когда значение текущей категории пользователя может меняться, в данном случае равно 16. Данные представлены в качестве матрицы объектов-признаков размера  $1644 \times 36$ . Столбцы матрицы соответствуют признакам:



**Таблица 1** Точность классификации пользователя в одну из объективных категорий в каждый момент времени  $t$  разными классификаторами

Классификатор	$t$													
	1	2	3	4	5	7	8	9	10	11	13	14	15	16
Байесовский классификатор	0,76	0,84	0,88	0,82	0,89	0,65	0,68	0,73	0,75	0,79	0,59	0,61	0,64	0,67
Логистическая регрессия	0,85	0,92	0,85	0,90	0,92	0,78	0,80	0,80	0,81	0,80	0,79	0,86	0,89	0,92
KNN	0,83	0,91	0,83	0,89	0,91	0,68	0,75	0,75	0,82	0,82	0,76	0,86	0,82	0,85
Дерево решений	0,77	0,83	0,87	0,87	0,87	0,66	0,77	0,83	0,80	0,79	0,67	0,77	0,80	0,84
Случайный лес	0,78	0,91	0,86	0,90	0,93	0,81	0,86	0,87	0,86	0,83	0,85	0,92	0,88	0,94
Градиентный бустинг	0,83	0,90	0,86	0,90	0,92	0,82	0,84	0,85	0,85	0,81	0,83	0,90	0,88	0,90

- оценки  $k_{t_i}$ ,  $i = 5, 10, 15$ , за очные контрольные работы, проводившиеся в моменты времени  $t_i$  (столбцы 1–3);
- нормированные суммарные сложности правильно решенных заданий тестов в ЭУ, выполненных в моменты времени  $t_i$  (столбцы 4–19);
- доля пропущенных пользователем очных занятий, прошедших в период до момента времени  $t_i$  (столбцы 20–36).

Каждая строка является признаковым описанием одного пользователя, всего пользователей в выборке 1644; 70% взято в качестве обучающей выборки; 30% — тестовая выборка.

В табл. 1 приведены точности прогнозирования объективной категории пользователя, формируемой на следующем за моментом времени  $t$  очном контрольном мероприятии, для каждого из рассматриваемых классификаторов. Также в табл. 1 приведены результаты байесовского классификатора, описанного в [4].

Точность классификации — доля правильно определенных объективных категорий пользователей классификатором на тестовой выборке.

Все классификаторы, рассматриваемые в данной работе, показали лучший результат по сравнению с байесовским классификатором, наилучшие прогнозы дали случайный лес и градиентный бустинг. Также можно заметить, что после каждой контрольной работы точность классификации с помощью байесовского метода довольно сильно уменьшается, у остальных методов не было такой проблемы.

Стоит отметить, что, в отличие от байесовского классификатора, другие исследуемые методы требуют задания внутренних параметров. Так, число ближайших соседей в методе KNN выбиралось простым перебором — число увеличивалось до тех пор, пока не наблюдалось переобучение. Параметры других классификаторов подбирались схожим образом. И хотя оптимизация гиперпара-



метров классификатора в общем случае задача нетривиальная, сама возможность настройки используемого метода позволяет добиваться лучшей точности классификации.

Применение байесовского классификатора потребовало меньших вычислительных и временных затрат, что стало бы плюсом при решении задач с большим набором исходных данных.

Далее все классификаторы, кроме байесовского, были объединены в один классификатор. Новый классификатор предсказывает объективную категорию по принципу большинства.

Обозначим  $a_k = \{2, 3, 4, 5\}$  — ответ  $k$ -го классификатора (прогнозируемая объективная категория студента), где  $k = 1, \dots, 5$  — номер классификатора. Тогда результирующий алгоритм для каждого студента будет иметь вид:

$$a(\cdot) = \arg \max_{y_{t_i}} \sum_{a_k} I(a_k = y_{t_i}),$$

где  $y_{t_i} \in Y = \{0, 2, 3, 4, 5\}$  — возможная объективная категория студента в момент времени  $t_i$  очередной очной контрольной работы, следующей за моментом времени проведения классификации.

Объединение различных алгоритмов позволяет снизить риск переобучения, а также неправильной интерпретации результатов какой-либо одной отдельно взятой модели. В то же время такой подход значительно увеличивает время работы классификатора. Также можно попробовать улучшить этот метод, добавив весовые коэффициенты составляющим алгоритмам, если, например, какому-то из них доверяем больше остальных. Результаты работы описанного классификатора приведены в табл. 2.

Точность объединенного классификатора несколько ниже, чем у случайного леса и градиентного бустинга, однако такая модель более стабильна.

**Таблица 2** Точность классификации пользователя в одну из объективных категорий в каждый момент времени  $t$  объединенным классификатором

$t$	Точность
1	0,78
2	0,95
3	0,91
4	0,91
5	0,91
7	0,75
8	0,78
9	0,81
10	0,90
11	0,82
13	0,89
14	0,90
15	0,84
16	0,91

## 4 Заключение

В ходе данной работы было продолжено исследование модели формирования индивидуальной траектории пользователя СДО [4] при смешанной форме ведения образовательной деятельности с организацией самостоятельной работы обучаемых с помощью СДО.

По сравнению с [4] было рассмотрено другое признаковое пространство. Для прогнозирования объективной категории пользователей с учетом нового представления данных использовались следующие классификаторы: логистическая

регрессия, метод ближайших соседей (KNN), дерево решений, случайный лес, градиентный бустинг. Данные классификаторы показали бóльшую точность классификации по сравнению с байесовским классификатором. Кроме того, было обнаружено, что точность классификации байесовским методом становится хуже после каждой контрольной работы, что несвойственно остальным методам. Это объясняется тем, что другие методы эффективнее используют новое признаковое пространство. Байесовский классификатор использует свертку, усредняя и складывая значения признаков, что приводит к потере информации о последовательном изменении этих значений и других закономерностях в данных. При решении задачи классификации пользователя СДО в одну из объективных категорий рейтинга эта информация играет важную роль, так как может, например, показывать последовательное улучшение либо ухудшение результатов пользователя в процессе обучения. Таким образом, с каждой новой контрольной работой теряется все больший объем информации, что снижает точность классификации.

Также все классификаторы, кроме байесовского, были объединены в один классификатор. Точность нового классификатора оказалась немного хуже, чем у случайного леса и градиентного бустинга, однако такая модель более устойчива к переобучению, так как учитывает результаты работы нескольких различных классификаторов.

## Литература

1. *Куравский Л. С., Мармалюк П. А., Алхимов В. И., Юрьев Г. А.* Новый подход к построению интеллектуальных и компетентностных тестов // Моделирование и анализ данных, 2013. № 1. С. 4–28.
2. *Kuravsky L. S., Margolis A. A., Marmalyuk P. A., Panfilova A. S., Yuryev G. A., Dumin P. N.* A probabilistic model of adaptive training // Applied Mathematical Sciences, 2016. Vol. 10. No. 48. P. 2369–2380. doi: 10.12988/ams.2016.65168.
3. *Чумакова Е. В., Корнеев Д. Г., Гаспарян М. С.* Разработка метода адаптивного тестирования на основе нейротехнологий // Открытое образование, 2022. Т. 26. № 2. С. 4–13. doi: 10.21686/1818-4243-2022-2-4-13.
4. *Босов А. В., Мартюшова Я. Г., Наумов А. В., Сапунова А. П.* Байесовский подход к построению индивидуальной траектории пользователя в системе дистанционного обучения // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 86–93. doi: 10.14357/19922264200313.
5. *Босов А. В.* Применение самоорганизующихся нейронных сетей к процессу формирования индивидуальной траектории обучения // Информатика и её применения, 2022. Т. 16. Вып. 3. С. 7–15. doi: 10.14357/19922264220302.
6. *Bosov A. V.* Adaptation of Kohonen's self-organizing map to the task of constructing an individual user trajectory in an e-learning system // Data science and algorithms in systems / Eds. R. Silhavy, P. Silhavy, Z. Prokopova. — Lecture notes in networks and systems ser. — Cham: Springer, 2022. Vol. 597. P. 554–564. doi: 10.1007/978-3-031-21438-7\_44.

7. *Rasch G.* Probabilistic models for some intelligence and attainment tests. — Chicago, IL, USA: The University of Chicago Press, 1980. 199 p.
8. *Kohonen T.* Self-organizing maps. — 3rd ed. — Berlin, Heidelberg: Springer, 2001. 501 p.
9. *Ренкина Н. Г.* Прогнозирование успешности образования студентов технических направлений подготовки с использованием искусственных нейронных сетей // Альманах мировой науки, 2016. № 5-1(8). С. 92–95.
10. *Казаченок В. В.* Применение нейронных сетей в обучении // Информатика и образование, 2020. № 2(311). С. 41–47. doi: 10.32517/0234-0453-2020-35-2-41-47.
11. *Okewu E., Adewole P., Misra S., Maskeliunas R., Damasevicius R.* Artificial neural networks for educational data mining in higher education: A systematic literature review // Appl. Artif. Intell., 2021. Vol. 35. Iss. 13. P. 983–1021. doi: 10.1080/08839514.2021.1922847.
12. *Шамсутдинова Т. М.* Проблемы и перспективы применения нейронных сетей в сфере образования // Открытое образование, 2022. Т. 26. № 6. С. 4–10. doi: 10.21686/1818-4243-2022-6-4-10.
13. *Luo Q., Yang J.* The artificial intelligence and neural network in teaching // Comput. Intel. Neurosc., June 10, 2022. Art. 1778562. 11 p. doi: 10.1155/2022/1778562.
14. *Наумов А. В., Джумурат А. С., Иноземцев А. О.* Система дистанционного обучения математическим дисциплинам CLASS.NET // Вестник компьютерных и информационных технологий, 2014. № 10. С. 36–44. doi: 10.14489/vkit.2014.010.pp.036-044.
15. СДО МАИ CLASS.NET. <http://www.distance.mai.ru>.
16. *Наумов А. В., Иноземцев А. О.* Алгоритм формирования индивидуальных заданий в системах дистанционного обучения // Вестник компьютерных и информационных технологий, 2013. № 6. С. 46–51.
17. *Мартюшова Я. Г., Лыкова Н. М.* Организация рефлексивно-оценочной деятельности студентов университетов средствами электронного учебника // Психолого-педагогические исследования, 2018. Т. 10. № 2. С. 125–134. doi: 10.17759/psyedu.2018100211.

*Поступила в редакцию 06.02.23*

---

---

## METHODS OF CLASSIFYING THE DISTANCE LEARNING SYSTEM USERS IN THE MODEL OF CONSTRUCTING THEIR PERSONALIZED LEARNING STRATEGIES

*Ya. G. Martyushova, T. A. Mineyeva, and A. V. Naumov*

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

**Abstract:** The article examines the problem of adaptation of the distance learning system to the contingent of users by constructing personalized learning strategies using their previous tests results. The main part of the suggested model is classifying users by various academic progress criteria. Comparative analysis of results of applying different classifiers for this purpose is presented.

The following types of classifiers were used: Bayes classifier, logistic regression, k-nearest neighbors algorithm, decision tree, random forest, boosting, and bootstrap aggregating classifier that uses a majority vote as the voting scheme. The article presents the results of a numerical experiment using the data on the work of MAI distance learning system CLASS.NET.

**Keywords:** distance learning system; machine learning methods; adaptive system; personalized online education learning strategies

**DOI:** 10.14357/08696527230107

## Acknowledgments

This work was partially supported by the Russian Science Foundation (grant No. 22-28-00588, <https://rscf.ru/project/22-28-00588>).

## References

1. Kuravsky, L. S., P. A. Marmalyuk, V. I. Alkhimov, and G. A. Yuryev. 2013. Novyy podkhod k postroeniyu intellektual'nykh i kompetentnostnykh testov [A new approach to constructing intellectual and competence-based tests]. *Modelirovanie i analiz dannykh* [Modelling and Data Analysis] 1:4–28.
2. Kuravsky, L. S., A. A. Margolis, P. A. Marmalyuk, A. S. Panfilova, G. A. Yuryev, and P. N. Dumin. 2016. A probabilistic model of adaptive training. *Applied Mathematical Sciences* 10(48):2369–2380. doi: 10.12988/ams.2016.65168.
3. Chumakova, E. V., D. G. Korneev, and M. S. Gasparian. 2022. Razrabotka metoda adaptivnogo testirovaniya na osnove neyrotekhnologiy [Development of adaptive testing method based on neurotechnologies]. *Otkrytoe obrazovanie* [Open Education] 26(2):4–13. doi: 10.21686/1818-4243-2022-2-4-13.
4. Bosov, A. V., Ya. G. Martyushova, A. V. Naumov, and A. P. Sapunova. 2020. Bayesovskiy podkhod k postroeniyu individual'noy traektorii pol'zovatelya v sisteme distantsionnogo obucheniya [Bayesian approach to the construction of an individual user trajectory in the system of distance learning]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):86–93. doi: 10.14357/19922264200313.
5. Bosov, A. V. 2022. Primenenie samoorganizuyushchikhsya neyronnykh setey k protsessu formirovaniya individual'noy traektorii obucheniya [Application of self-organizing neural networks to the process of forming an individual learning path]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3):7–15. doi: 10.14357/19922264220302.
6. Bosov, A. V. 2022. Adaptation of Kohonen's self-organizing map to the task of constructing an individual user trajectory in an e-learning system. *Data science and algorithms in systems*. Eds. R. Silhavy, P. Silhavy, and Z. Prokopova. Lecture notes in networks and systems ser. Cham: Springer. 597:554–564. doi: 10.1007/978-3-031-21438-7\_44.
7. Rasch, G. 1980. *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: The University of Chicago Press. 199 p.
8. Kohonen, T. 2001. *Self-organizing maps*. 3rd ed. Berlin, Heidelberg: Springer. 501 p.
9. Repkina, N. G. 2016. Prognozirovanie uspekhov obrazovaniya studentov tekhnicheskikh napravleniy podgotovki s ispol'zovaniem iskusstvennykh neyronnykh setey

- [Prediction of success of educating students of technical arias using artificial neural networks]. *Al'manakh mirovoy nauki* [Almanac of World Science] 5-1(8):92–95.
10. Kazachonak, V. V. 2020. Primenenie neyronnykh setey v obuchenii [Application of neural networks in training]. *Informatika i obrazovanie* [Informatics and Education] 2(311):41–47. doi: 10.32517/0234-0453-2020-35-2-41-47.
  11. Okewu, E., P. Adewole, S. Misra, R. Maskeliunas, and R. Damasevicius. 2021. Artificial neural networks for educational data mining in higher education: A systematic literature review. *Appl. Artif. Intell.* 35(13):983–1021. doi: 10.1080/08839514.2021.1922847.
  12. Shamsutdinova, T. M. 2022. Problemy i perspektivy primeneniya neyronnykh setey v sfere obrazovaniya [Problems and prospects for the application of neural networks for the sphere of education]. *Otkrytoe obrazovanie* [Open Education] 26(6):4–10. doi: 10.21686/1818-4243-2022-6-4-10.
  13. Luo, Q., and J. Yang. June 10, 2022. The artificial intelligence and neural network in teaching. *Comput. Intel. Neurosc.* 1778562. 11 p. doi: 10.1155/2022/1778562.
  14. Naumov, A. V., A. S. Dzhumurat, and A. O. Inozemtsev. 2014. Sistema distantsionno-go obucheniya matematicheskim distsiplinam CLASS.NET [Distance learning system for mathematical disciplines CLASS.NET]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies] 10:36–44. doi: 10.14489/vkit.2014.010.pp.036-044.
  15. SDO MAI CLASS.NET [MAI distance learning system CLASS.NET]. Available at: <http://www.distance.mai.ru/> (accessed January 10, 2023)
  16. Naumov, A. V., and A. O. Inozemtsev. 2013. Algoritm formirovaniya individual'nykh zadaniy v sistemakh distantsionnogo obucheniya [The algorithm for generating the individual tasks in distance learning]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies] 6:46–51.
  17. Martyushova, Ya. G., and N. M. Lykova. 2018. Organizatsiya refleksivno-otsenочноy deyatel'nosti studentov universitetov sredstvami elektronnoy uchebnika [Organization of reflexive-evaluative activity of university students by using the learning management system]. *Psikhologo-pedagogicheskie issledovaniya* [Psychological-Educational Studies] 10(2):125–134. doi: 10.17759/psyedu.2018100211.

*Received February 6, 2023*

## Contributors

**Martyushova Yanina G.** (b. 1970) — Candidate of Science (PhD) in pedagogy, associate professor, Probability Theory and Computer Simulations Department, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; ma1554@mail.ru

**Mineeva Tatyana A.** (b. 2002) — bachelor student, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; mineeva.t2015@yandex.ru

**Naumov Andrey V.** (b. 1966) — Doctor of Science in physics and mathematics, professor, Probability Theory and Computer Simulations Department, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; naumovav@mail.ru

## ЭФФЕКТИВНЫЕ ВЫЧИСЛЕНИЯ ПРИ ФАКТОРИЗАЦИИ МАТРИЧНЫХ ДАННЫХ С ПРОПУСКАМИ

*М. П. Кривенко*<sup>1</sup>

**Аннотация:** Работа посвящена эффективной реализации факторизации матрицы с отсутствующими компонентами в произведение двух матриц более низкого ранга. Задача оценивания параметров принятой модели данных решается путем многомерной оптимизации. Большие размеры матриц и векторов, входящих в итерационные алгоритмы, порождают на практике проклятие размерности. Предлагается кардинально снизить сложность матричных операций путем представления их в блочно-диагональном виде. Обосновывается возможность приведения обрабатываемых матриц к блочно-диагональному виду, описываются правила поблочного сингулярного разложения матриц. Преимущества поблочной обработки иллюстрируются на примере факторизации матрицы данных различных размеров и с различными вероятностями пропуска компонента. Время оценивания параметров может сокращаться на несколько порядков по сравнению с обработкой матриц в обычном представлении.

**Ключевые слова:** понижающая ранг аппроксимация матрицы; сингулярное разложение; пропущенные данные; ALS-алгоритм; блочно-диагональное представление матрицы

**DOI:** 10.14357/08696527230108

### 1 Введение

Факторизация матриц данных хорошо зарекомендовала себя как метод снижения размерности. Множество его приложений включает обработку изображений, сжатие передаваемой информации, разведочный анализ данных, визуализацию, распознавание образов и прогнозирование временных рядов. Трудности при факторизации начинают проявляться при возникновении пропусков в данных, причем наилучших результатов удается достичь с помощью методов обработки наблюдаемых и пропущенных данных в совокупности, которые требуют разработки оригинальных процедур. С общей характеристикой этой проблемы можно ознакомиться, например, в [1]. В [2] используемые методы факторизации матриц данных с пропусками дополнены деталями, выявляющими и исправляющими ошибки в доступных публикациях. Обращено внимание на необходимость исследования существующих и создания новых алгоритмов, учитывающих специфические особенности обрабатываемых матриц.

---

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

Модель факторизации  $(m \times n)$ -матрицы наблюдений  $\mathbf{Y}$  — это представление ее в виде  $\tilde{\mathbf{Y}} = \mathbf{U}\mathbf{V}^T$ , где  $\mathbf{U}$  —  $(m \times r)$ -матрица;  $\mathbf{V}$  —  $(n \times r)$ -матрица. Наличие пропусков в данных отражается в  $(m \times n)$ -матрице  $\mathbf{H}$ :

$$h_{ij} = \begin{cases} 1, & y_{ij} \text{ присутствует;} \\ 0, & y_{ij} \text{ пропущено.} \end{cases}$$

Для суммы всех ее элементов примем обозначение  $p$  (число присутствующих элементов матрицы наблюдений  $\mathbf{Y}$ ).

**Задача построения  $\tilde{\mathbf{Y}}$**  формулируется как минимизация целевой функции

$$\varphi(\mathbf{U}, \mathbf{V}) = \left\| \mathbf{H} \odot (\mathbf{Y} - \tilde{\mathbf{Y}}) \right\|_F^2 \rightarrow \min_{\mathbf{U}, \mathbf{V}}.$$

Перепишем целевую функцию в более удобном для аналитических преобразований виде, исключив  $\mathbf{H}$ . Построчная запись матрицы  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]^T$ , где  $\mathbf{u}_i$  суть  $r$ -векторы, даст  $mr$ -вектор  $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_m^T]^T$ . Также для матрицы  $\mathbf{V}$  определим  $nr$ -вектор  $\mathbf{v}$ . В результате  $\varphi(\mathbf{U}, \mathbf{V})$  можно переписать как

$$\varphi(\mathbf{U}, \mathbf{V}) \equiv \varphi(\mathbf{u}, \mathbf{v}) = |\mathbf{F}\mathbf{u} - \mathbf{y}|^2 = |\mathbf{G}\mathbf{v} - \mathbf{y}|^2, \quad (1)$$

где  $p$ -вектор  $\mathbf{y}$  формируется из соответствующих элементов  $\mathbf{Y}$ , а  $(p \times mr)$ -матрица  $\mathbf{F}$  — из векторов  $\mathbf{v}_i$  и  $(p \times nr)$ -матрица  $\mathbf{G}$  — из векторов  $\mathbf{u}_i$  согласно только значениям  $h_{ij} = 1$ . Смысл (1) в том, что параметры модели (векторы  $\mathbf{u}$  и  $\mathbf{v}$ ) разделены в двух эквивалентных представлениях целевой функции.

Для нахождения минимума (1) принят альтернирующий алгоритм наименьших квадратов (Alternating Least Squares, ALS). Он относится к итерационным и подразумевает задание начального значения вектора  $\mathbf{v}$ , что порождает матрицу  $\mathbf{F}$ . Далее происходит пересчет по формулам

$$\mathbf{u} = \left( \mathbf{F}^T \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{y}, \quad (2)$$

что дает новое значение  $\mathbf{G}$ ,

$$\mathbf{v} = \left( \mathbf{G}^T \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{y}, \quad (3)$$

что дает новое значение  $\mathbf{F}$  и завершает очередной шаг итерации.

Обычно ключевым считается предположение о том, что матрицы  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{F}$  и  $\mathbf{G}$  — полного ранга. Проведенные эксперименты показали, что это не так. Выявленные контрпримеры поведения рангов основополагающих матриц объясняют неустойчивое поведение привычных алгоритмов оценивания параметров модели и приводят к необходимости их коррекции. Если кратко, то изменения свелись к контролю рангов обрабатываемых матриц и использованию псевдообращения Мура–Пенроуза (Moore–Penrose Inverse, MPI).



Сведение задачи факторизация матрицы данных с пропусками к оптимизации (1) порождает проблемы многомерности: достаточно обратить внимание на размеры входящих в (1) векторов и матриц, а также понять, что для исследования свойств получающихся решений придется использовать методы статистических испытаний.

Обращает на себя внимание разреженность матриц  $\mathbf{F}$  и  $\mathbf{G}$ : доля нулевых элементов не менее  $(m - 1)/m$ , а также их специфический блочный характер (в случае  $\mathbf{F}$  достаточно изобразить ее схематически, а вот для  $\mathbf{G}$  это не очевидно). Как следствие, возникают предпосылки для создания эффективных вычислительных алгоритмов.

## 2 Блочно-диагональные матрицы. Особенности операций с ними

Согласно общепринятому определению, блочно-диагональная матрица  $\mathbf{A}$  имеет вид:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{qq} \end{pmatrix},$$

где  $\mathbf{A}_{ii}$  — квадратные матрицы не обязательно одинакового размера [3]. Расширим это определение, допустив прямоугольность блоков. Тогда если  $\mathbf{A}$  имеет размеры  $(s \times t)$ , а  $\mathbf{A}_{ii}$  —  $(s_i \times t_i)$ , то

$$s_1 + \cdots + s_q = s; \quad t_1 + \cdots + t_q = t.$$

Для введенных в рассмотрение матриц далее кратко обосновываются правила работы с ними.

В первую очередь понадобится представление для временной сложности  $T$  выполнения операции умножения матриц. Обычное умножение  $(s \times t)$ -матрицы  $\mathbf{A}$  на  $(t \times u)$ -матрицу  $\mathbf{B}$  дает  $T_0(\mathbf{AB}) = stu$ . Если же матрицы  $\mathbf{A}$  и  $\mathbf{B}$  суть блочно-диагональные с  $q$  блоками, имеющими надлежащие размеры —  $(s_i \times t_i)$  и  $(t_i \times u_i)$ , то  $\mathbf{AB}$  — также блочно-диагональная с соответствующими размерами и  $T_b(\mathbf{AB}) = \sum_{i=1}^q s_i t_i u_i$ . В случае, когда вместо  $\mathbf{B}$  фигурирует  $t$ -вектор,  $T_b(\mathbf{Ab}) = \sum_{i=1}^q s_i t_i$ .

Согласно правилам транспонирования и перемножения блочных матриц получаем

$$\mathbf{A}^T \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11}^T \mathbf{A}_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_{qq}^T \mathbf{A}_{qq} \end{pmatrix}. \quad (4)$$

Для обращения блочно-диагональных матриц  $\mathbf{A}^T \mathbf{A}$  в виде правой части (4) можно использовать предварительно преобразованные соотношения для  $(2 \times 2)$



блочных матриц [4]. Но проще, подобно правилам построения (4), сначала предположить, что

$$\left(\mathbf{A}^T \mathbf{A}\right)^+ = \begin{pmatrix} \left(\mathbf{A}_1^T \mathbf{A}_1\right)^+ & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\mathbf{A}_{qq}^T \mathbf{A}_{qq}\right)^+ \end{pmatrix}, \quad (5)$$

а далее проверить выполнимость всех соотношений из определения МРІ.

Возможность замены операций со всей матрицей на поблочную обработку желательно обосновать и при ее факторизации (например, в целях реализации последовательности операций с матрицами или вычисления ранга матрицы). В первую очередь речь идет о сингулярном разложении (Singular Value Decomposition, SVD), которое далее строится через спектральное разложение (Spectral Decomposition, SD) для симметричной вещественнозначной  $(s \times s)$  блочно-диагональной матрицы  $\mathbf{C} = \mathbf{A}^T \mathbf{A}$ . Для каждого из  $q$  блоков размера  $(s_i \times s_i)$  имеем SD вида

$$\mathbf{C}_{ii} = \dot{\mathbf{V}}_{ii} \dot{\Sigma}_{ii}^2 \dot{\mathbf{V}}_{ii}^T.$$

Образует блочно-диагональные матрицы:  $\dot{\mathbf{V}}$  размера  $(t \times t)$  из  $\dot{\mathbf{V}}_{ii}$  и  $\dot{\Sigma}^2$  размера  $(t \times t)$  из  $\dot{\Sigma}_{ii}^2$ , где  $t = s_1 + \cdots + s_q$ . Тогда согласно правилам оперирования с блочными матрицами и свойствам SD получаем, что

$$\mathbf{C} = \dot{\mathbf{V}} \dot{\Sigma}^2 \dot{\mathbf{V}}^T \quad \text{и} \quad \dot{\mathbf{V}} \dot{\mathbf{V}}^T = \mathbf{I}. \quad (6)$$

Этого разложения уже достаточно для обращения матрицы или подсчета ранга. Но в общем случае ненулевые элементы матрицы  $\dot{\Sigma}^2$  — собственные значения отдельных блоков, хотя и образуют диагональ, но не упорядочены в совокупности, поэтому формально (6) не является SD.

Докажем, что элементы главной диагонали диагональной матрицы с помощью перестановочной матрицы можно поменять между собой. Для этого обратимся к известному результату: для произвольных  $(t \times t)$ -матрицы  $\mathbf{B}$  и  $(t \times t)$ -матрицы перестановки  $\mathbf{P}$  диагональные элементы  $\mathbf{PBP}^T$  те же самые (возможно, переставленные), что и для  $\mathbf{B}$  [5]. Осталось доказать, что в  $\mathbf{PBP}^T$  вне диагонали стоят только недиагональные элементы  $\mathbf{B}$ . Используя представление  $\mathbf{P}$  в виде целочисленного вектора  $\mathbf{p}$  ( $i$ -я координата  $\mathbf{p}_i$  говорит, что при умножении  $\mathbf{B}$  слева на  $\mathbf{P}$  строка с номером  $\mathbf{p}_i$  перемещается на место  $i$ -й строки), получаем для всех  $j$ :  $(\mathbf{PB})_{i,j} = \mathbf{B}_{\mathbf{p}_i,j}$ . Умножение справа приводит к перестановке столбцов ( $j$ -й столбец должен переместиться в  $i$ -й, если в матрице соответствующей перестановки в  $i$ -й строке стоит 1) при матрице перестановки  $\mathbf{P}^T$ , т. е. для того же вектора  $\mathbf{p}$ , но с  $j$ -й координатой. В итоге имеем

$$\left(\mathbf{PBP}^T\right)_{i,j} = \left(\mathbf{B}\right)_{\mathbf{p}_i,\mathbf{p}_j}. \quad (7)$$

Если обозначить обратную к  $\mathbf{p}$  перестановку как  $\mathbf{p}^{-1}$ , то получаем

$$(\mathbf{B})_{ij} = \left( \mathbf{PBP}^T \right)_{p_i^{-1}, p_j^{-1}}.$$

Из полученных соотношений видно, что при  $i = j$  (диагональные элементы матриц) наборы всех различных значений элементов матриц  $\mathbf{B}$  и  $\mathbf{PBP}^T$  совпадают, а при  $i \neq j$  согласно (7) в случае диагональной матрицы  $\mathbf{B}$  верно, что  $(\mathbf{PBP}^T)_{i,j} = 0$ .

Таким образом, если неупорядоченность диагональных элементов  $\Sigma^2$  препятствует использованию блочного SD, то можно внести коррективы и получить привычное разложение, для чего необходимо:

- вычленив в  $\Sigma^2$  все ненулевые элементы, зафиксировав их положение в матрице;
- упорядочить полученные значения по убыванию, сохранив ссылки на их исходные положения в матрице;
- сформировать матрицу перестановок  $\mathbf{P}$  для строк и столбцов матрицы  $\Sigma^2$ ;
- получить диагональную матрицу  $\tilde{\Sigma}^2 = \mathbf{P}\Sigma^2\mathbf{P}^T$  с упорядоченными по убыванию числами на главной диагонали.

В результате имеем

$$\mathbf{A}^T \mathbf{A} = \mathbf{C} = \dot{\mathbf{V}}\Sigma^2\dot{\mathbf{V}}^T = \dot{\mathbf{V}}\mathbf{P}^T\tilde{\Sigma}^2\mathbf{P}\dot{\mathbf{V}}^T \equiv \ddot{\mathbf{V}}\tilde{\Sigma}^2\ddot{\mathbf{V}}^T.$$

При этом с учетом ортогональности перестановочной матрицы

$$\ddot{\mathbf{V}}\ddot{\mathbf{V}}^T = \left( \dot{\mathbf{V}}\mathbf{P}^T \right) \left( \dot{\mathbf{V}}\mathbf{P}^T \right)^T = \dot{\mathbf{V}}\mathbf{P}^T\mathbf{P}\dot{\mathbf{V}}^T = \mathbf{I}.$$

Таким образом, получено SD матрицы  $\mathbf{A}^T \mathbf{A}$ . Заметим, что в общем случае матрица  $\dot{\mathbf{V}}$  после умножения на перестановочную матрицу перестает быть блочно-диагональной.

Учитывая связь между SD и SVD в случае произвольной прямоугольной матрицы  $\mathbf{A}$ , из представления  $\mathbf{A}^T \mathbf{A} = \dot{\mathbf{V}}\tilde{\Sigma}^2\dot{\mathbf{V}}^T$  можно получить  $\ddot{\mathbf{V}}$  и  $\tilde{\Sigma}^2$ , а затем найти  $\ddot{\mathbf{U}} = \mathbf{A}\tilde{\Sigma}^{-1}$ , что приводит к представлению  $\mathbf{A} = \ddot{\mathbf{U}}\tilde{\Sigma}\ddot{\mathbf{V}}^T$ , т. е. к SVD.

Идеи перестановочной эквивалентности позволяют получить SVD прямым путем, опираясь на представления отдельных блоков

$$\mathbf{A}_{ii} = \bar{\mathbf{U}}_{ii}\bar{\Sigma}_{ii}\bar{\mathbf{V}}_{ii}^T,$$

где размеры матриц соответственно равны:  $(s_i \times t_i)$ ,  $(s_i \times s_i)$ ,  $(s_i \times t_i)$  и  $(t_i \times t_i)$ .

Образует блочно-диагональные матрицы:  $\bar{\mathbf{U}}$  размера  $(s \times s)$  из  $\bar{\mathbf{U}}_{ii}$ ,  $\bar{\mathbf{V}}^T$  размера  $(t \times t)$  из  $\bar{\mathbf{V}}_{ii}^T$ ,  $\bar{\mathbf{\Sigma}}$  размера  $(s \times t)$  из  $\bar{\Sigma}_{ii}$ , где, как и ранее,  $s$  и  $t$  — суммы соответствующих одноименных элементов. Тогда согласно правилам оперирования с блочными матрицами получаем, что  $\mathbf{A} = \bar{\mathbf{U}} \bar{\mathbf{\Sigma}} \bar{\mathbf{V}}^T$  и  $\bar{\mathbf{U}} \bar{\mathbf{U}}^T = \mathbf{I}$ ,  $\bar{\mathbf{V}} \bar{\mathbf{V}}^T = \mathbf{I}$ . Но в общем случае матрица  $\bar{\mathbf{\Sigma}}$  не диагональная, а ее элементы — сингулярные значения отдельных блоков — никак не упорядочены. Если последнее препятствует использованию поблочного SVD, то можно внести коррективы и получить привычное разложение, для чего:

- вычленив в  $\bar{\mathbf{\Sigma}}$  все ненулевые элементы (тот факт, что в каждой строке и каждом столбце  $\bar{\mathbf{\Sigma}}$  либо все 0, либо только один ненулевой элемент, остается при блочном разложении в силе), зафиксировав их положение в матрице;
- упорядочить полученные значения по убыванию, сохранив ссылки на их исходные положения;
- сформировать матрицы перестановок  $\mathbf{P}_L$  для строк и  $\mathbf{P}_R$  для столбцов матрицы  $\bar{\mathbf{\Sigma}}$ , не забывая о полностью нулевых строках и столбцах  $\bar{\mathbf{\Sigma}}$ ;
- получить матрицу  $\bar{\bar{\mathbf{\Sigma}}} = \mathbf{P}_L \bar{\mathbf{\Sigma}} \mathbf{P}_R$  с упорядоченными по убыванию числами на главной диагонали.

В результате имеем:

$$\mathbf{A} = \bar{\mathbf{U}} \bar{\mathbf{\Sigma}} \bar{\mathbf{V}}^T = \bar{\mathbf{U}} \mathbf{P}_L^T \bar{\bar{\mathbf{\Sigma}}} \mathbf{P}_R^T \bar{\mathbf{V}}^T = \bar{\bar{\mathbf{U}}} \bar{\bar{\mathbf{\Sigma}}} \bar{\bar{\mathbf{V}}}^T.$$

Заметим, что блочная структура  $\bar{\bar{\mathbf{U}}}$  и  $\bar{\bar{\mathbf{V}}}^T$  в общем случае нарушена.

Учитывая ортогональность перестановочных матриц, получаем

$$\bar{\bar{\mathbf{U}}} \bar{\bar{\mathbf{U}}}^T = (\bar{\mathbf{U}} \mathbf{P}_L^T) (\bar{\mathbf{U}} \mathbf{P}_L^T)^T = \bar{\mathbf{U}} \mathbf{P}_L^T \mathbf{P}_L \bar{\mathbf{U}}^T = \mathbf{I};$$

$$\bar{\bar{\mathbf{V}}} \bar{\bar{\mathbf{V}}}^T = (\bar{\mathbf{V}} \mathbf{P}_R) (\bar{\mathbf{V}} \mathbf{P}_R)^T = \bar{\mathbf{V}} \mathbf{P}_R \mathbf{P}_R^T \bar{\mathbf{V}}^T = \mathbf{I}.$$

Таким образом, получено SVD для матрицы  $\mathbf{A}$ .

Важно помнить, что сингулярные числа уникальны, а матрицы сингулярных векторов нет. Более того, реально существуют иные версии SVD [6, разд. 16.3], фактически в данной работе к ним добавлена еще одна — поблочная.

### 3 Реализация ALS-алгоритма при блочном представлении $\mathbf{F}$ и $\mathbf{G}$

**Матрица  $\mathbf{F}$ .** На основе  $i$ -й строки матрицы  $\mathbf{H}$ ,  $i = 1, \dots, m$ , в  $(p \times mr)$ -матрице  $\mathbf{F}$  возникает блок  $\mathbf{F}_i$  размера  $(s_i \times r)$ , где  $s_i$  — число единиц в этой строке, а сама матрица принимает блочно-диагональный вид. Тогда  $\mathbf{F}^T \mathbf{F}$  становится

также блочно-диагональной и ее обращение принимает вид (5). Таким образом, реализация (2) сводится к обработке двух блочно-диагональных матриц, каждая из которых есть набор  $m$  матриц.

Нагляднее можно описать формирование  $\mathbf{F}$  в виде алгоритма в предположении, что структура блочно-диагональной матрицы уже сформирована. Дополнительно введем для каждого блока индекс  $s_i$  текущей формируемой строки блока. Он носит специфичный характер, отражающий последовательное, строка за строкой, формирование блоков, а также обеспечивает возможность многократного использования в ходе итерационного процесса статически распределенной памяти под блок. В итоге получаем

### Алгоритм заполнения $\mathbf{F}$

```

for  $i := 1$  to  $m$  do  $s_i := 0$ ; // задание индексов формируемых строк блоков
for  $i_H := 1$  to  $m$  do // перебор строк  $\mathbf{H}$ 
  for  $j_H := 1$  to  $n$  do // перебор столбцов  $\mathbf{H}$ 
    if  $H[i_H, j_H] = 1$ 
      then begin // обработка присутствующего элемента данных
         $b = i_H$ ; // индекс блока (выделено для наглядности)
         $s_b := s_b + 1$ ; // переход к очередной формируемой строке  $\mathbf{F}_b$ 
        for  $i_u := 1$  to  $r$  do  $\mathbf{F}_b[s_b, i_u] := \mathbf{u}[i_H r + i_u]$ ;
      end;

```

**Матрица  $\mathbf{G}$**  не является блочно-диагональной, но с помощью перестановки ее строк этот недостаток можно устранить. Пусть  $\mathbf{P}_\mathbf{H}$  —  $(p \times p)$ -матрица перестановок, обеспечивающая получение блочно-диагональной  $(p \times nr)$ -матрицы  $\tilde{\mathbf{G}} = \mathbf{P}_\mathbf{H} \mathbf{G}$ , распадающейся на  $n$  блоков  $\tilde{\mathbf{G}}_1, \dots, \tilde{\mathbf{G}}_n$ . Возможность построения  $\mathbf{P}_\mathbf{H}$  следует из постоянной длины векторов  $\mathbf{u}_i$  и того, что любая строка  $\mathbf{G}$  включает  $n - 1$  нулевых векторов той же длины и один из векторов  $\mathbf{u}_1, \dots, \mathbf{u}_n$ .

Если вернуться к (3), то получаем:

$$\mathbf{G}^T \mathbf{G} = \left( \mathbf{P}_\mathbf{H}^T \tilde{\mathbf{G}} \right)^T \mathbf{P}_\mathbf{H}^T \tilde{\mathbf{G}} = \tilde{\mathbf{G}}^T \mathbf{P}_\mathbf{H} \mathbf{P}_\mathbf{H}^T \tilde{\mathbf{G}} = \tilde{\mathbf{G}}^T \tilde{\mathbf{G}};$$

$$\mathbf{G}^T \mathbf{y} = \left( \mathbf{P}_\mathbf{H}^T \tilde{\mathbf{G}} \right)^T \mathbf{y} = \tilde{\mathbf{G}}^T \mathbf{P}_\mathbf{H} \mathbf{y}.$$

В последнем соотношении обратим внимание на произведение  $\mathbf{P}_\mathbf{H} \mathbf{y}$ . Во-первых, перестановочную матрицу целесообразно представить в виде целочисленного  $p$ -вектора  $p_\mathbf{H}$ , содержащего в качестве  $i$ -го элемента  $p_\mathbf{H}(i)$  индекс столбца единственной «1» в  $i$ -й строке  $\mathbf{P}_\mathbf{H}$ . Тогда  $p_\mathbf{H}(i) = j$  эквивалентно  $\mathbf{P}_\mathbf{H}(i, j) = 1$  и означает при умножении  $\mathbf{P}_\mathbf{H}$  слева на согласованную матрицу перемещение ее  $j$ -й строки на место  $i$ -й. В частности, для  $\mathbf{P}_\mathbf{H} \mathbf{y}$  верно представление:

$$\left( (\mathbf{P}_\mathbf{H} \mathbf{y})_1, \dots, (\mathbf{P}_\mathbf{H} \mathbf{y})_p \right) = \left( y_{p_\mathbf{H}(1)}, \dots, y_{p_\mathbf{H}(p)} \right).$$

Во-вторых, (2) и (3) суть шаги итерационного процесса, в ходе которого вектор  $\mathbf{P}_H \mathbf{y}$  остается неизменным, т. е. может быть посчитан однократно.

При прежних предположениях и дополнительном обозначении  $f_i$  — индекс первой строки  $i$ -го блока в исходной матрице  $\mathbf{G}$ , получаем

**Алгоритм заполнения  $\tilde{\mathbf{G}}$ ,  $\mathbf{P}_H$  и  $p_H$**

```

 $\mathbf{P}_H := \mathbf{0}$ ; // обнуление матрицы перестановки
for  $i := 1$  to  $m$  do  $s_i := 0$ ; // задание индексов формируемых строк блоков
 $i_G := 1$ ; // задание индекса начальной строки  $\mathbf{G}$ 
for  $i_H := 1$  to  $m$  do // перебор строк  $\mathbf{H}$ 
  for  $j_H := 1$  to  $n$  do // перебор столбцов  $\mathbf{H}$ 
    if  $H[i_H, j_H] = 1$ 
      then begin // обработка присутствующего элемента данных
         $b = j_H$ ; // индекс блока (выделено для наглядности)
        for  $i_u := 1$  to  $r$  do  $\tilde{\mathbf{G}}_b[s_b, i_u] := \mathbf{u}[i_H r + i_u]$ ;
         $s_b := s_b + 1$ ; // переход к очередной формируемой строке  $\tilde{\mathbf{G}}_b$ 
         $\mathbf{P}_H[f_b + s_b, i_G] := 1$  //  $i_G$ -я строка перемещается в  $(f_b + s_b)$ -строку
         $p_H[f_b + s_b] := i_G$ ;
      end;
```

В итоге шаг (2) итерационного процесса останется прежним, а шаг (3) примет вид:

$$\mathbf{v} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y} = (\tilde{\mathbf{G}}^T \tilde{\mathbf{G}})^{-1} \tilde{\mathbf{G}}^T \mathbf{P}_H \mathbf{y}.$$

Требует правки и часть представления (1):

$$\varphi(\mathbf{u}, \mathbf{v}) = |\mathbf{G}\mathbf{v} - \mathbf{y}|^2 = |\mathbf{P}_H^T \tilde{\mathbf{G}}\mathbf{v} - \mathbf{y}|^2 = |\mathbf{P}_H^T \tilde{\mathbf{G}}\mathbf{v} - \mathbf{P}_H^T \mathbf{P}_H \mathbf{y}|^2 = |\tilde{\mathbf{G}}\mathbf{v} - \mathbf{P}_H \mathbf{y}|^2,$$

что позволяет в случае необходимости вычислять  $\varphi(\mathbf{u}, \mathbf{v})$  через введенные ранее объекты.

Рассмотрим теперь вычисление правой части (2) и напомним структуры входящих в нее элементов:

- $(\mathbf{F}^T \mathbf{F})^{-1}$  есть  $(mr \times mr)$ -матрица из  $m$  одинаковых блоков размера  $(r \times r)$ ;
- $\mathbf{F}^T$  есть  $(mr \times p)$ -матрица из  $m$  блоков размером  $(r \times s_1), \dots, (r \times s_m)$ ,  $s_1 + \dots + s_m = p$ ;
- $(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$  есть  $(mr \times p)$ -матрица из  $m$  блоков размером  $(r \times s_1), \dots, (r \times s_m)$ ;
- $\mathbf{y}$  есть  $p$ -вектор.

Теперь сравним объем вычислений для двух вариантов:

- (1) сначала  $\mathbf{A} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ , а затем  $\mathbf{A}\mathbf{y}$ , что даст  $T_{11} + T_{12}$ ;

(2) сначала  $\mathbf{B} = \mathbf{F}^T \mathbf{y}$ , а затем  $(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{B}$ , что даст  $T_{21} + T_{22}$ .

Получаем:

$$T_{11} = T_b \left( \left( (\mathbf{F}^T \mathbf{F})^{-1} \right) (\mathbf{F}^T) \right) = \sum_{i=1}^m r s_i r = r^2 p;$$

$$T_{12} = T_b \left( \left( (\mathbf{F}^T \mathbf{F})^{-1} \right) \mathbf{F}^T \right) (\mathbf{y}) = \sum_{i=1}^m r s_i = r p;$$

$$T_{21} = T_b (\mathbf{F}^T \mathbf{y}) = \sum_{i=1}^m r s_i = r p;$$

$$T_{22} = T_b \left( \left( (\mathbf{F}^T \mathbf{F})^{-1} \right) (\mathbf{F}^T \mathbf{y}) \right) = \sum_{i=1}^m r r = r^2 m.$$

Сравнивая различные по величине вклады в объем вычислений, получаем

$$\frac{T_{22}}{T_{11}} = \frac{m}{p},$$

что говорит о преимуществе 2-го варианта, так как  $p \geq m$  (в каждой строке матрицы исходных данных должен быть по крайней мере один элемент). Аналогичный результат справедлив и для матрицы  $\mathbf{G}$ . Таким образом, удастся получить дополнительные преимущества при обработке блочно-диагональных матриц за счет изменения порядка действий с ними.

Переход к блочному представлению позволяет сократить не только время вычислений, но и существенно снизить требования к объему используемой памяти. Если сравнивать эту характеристику для матрицы  $\mathbf{F}$  при ее обычном представлении и блочно-диагональном, то получим сокращение в  $p m r / (\sum_{i=1}^m s_i r) = m$  раз. Практически аналогичный результат верен и для  $\mathbf{G}$ .

Для иллюстрации реальных временных преимуществ поблочного оперирования при исполнении ALS-алгоритма рассмотрим факторизацию матрицы данных с фиксированным числом столбцов  $n = 20$  (число анализируемых признаков), увеличивающимся числом строк  $m = 50, 100, 200, 400$  (объемы выборок), различными вероятностями случайного пропуска элементов данных ( $P_m = 0,2; 0,4; 0,6$ ) и различными значениями числа факторов  $r = 5, 10, 15$ . Относительную эффективность поблочной обработки матриц оценим с помощью отношения  $E_r$  времени реализации ALS-алгоритма при обычном представлении матриц  $\mathbf{F}$  и  $\mathbf{G}$  ко времени реализации ALS-алгоритма при их блочно-диагональном представлении. Часть полученных результатов представлена в таблице.

Получившиеся оценки свидетельствуют о следующих моментах:

Относительная эффективность блочной обработки  $E_r$

m	$r = 5$		$r = 10$		$r = 15$	
	$P_m = 0,2$	$P_m = 0,6$	$P_m = 0,2$	$P_m = 0,6$	$P_m = 0,2$	$P_m = 0,6$
50	$3,3 \cdot 10^2$	$2,1 \cdot 10^2$	$4,9 \cdot 10^2$	$3,3 \cdot 10^2$	$4,6 \cdot 10^2$	$3,0 \cdot 10^2$
100	$1,1 \cdot 10^3$	$7,9 \cdot 10^2$	$1,8 \cdot 10^3$	$1,1 \cdot 10^3$	$1,9 \cdot 10^3$	$1,1 \cdot 10^3$
200	$3,5 \cdot 10^3$	$2,5 \cdot 10^3$	$6,1 \cdot 10^3$	$3,9 \cdot 10^3$	$7,5 \cdot 10^3$	$4,2 \cdot 10^3$
400	$8,8 \cdot 10^3$	$6,2 \cdot 10^3$	$1,4 \cdot 10^4$	$4,0 \cdot 10^3$	$1,8 \cdot 10^4$	$1,1 \cdot 10^4$

- безусловное преимущество использования блочно-диагонального представления матриц даже при необходимости дополнительных преобразований матриц к должному виду;
- явный рост преимущества с ростом  $m$ ;
- незначительная зависимость относительной эффективности от изменений  $P_m$  и  $r$ .

## 4 Заключение

Факторизация матрицы данных с пропусками сопряжена с вычислительными проблемами. Это экспоненциальный рост объема обрабатываемого материала и неизбежность реализации довольно сложных матричных операций. Кроме того, для практики крайне важен выбор значения  $r$ , для чего приходится обращаться к бутстреп-методу тиражирования исходных данных.

Полученные результаты показывают достоинства блочной обработки. Впечатляющие сами по себе, они могут стать отправной точкой для доводки вычислительных алгоритмов. В первую очередь речь идет об оптимизации процедур резервирования и активизации памяти, о выделении общих матричных операций и перестройке последовательности действий, о проработке возможности использования блочной версии SVD в чистом виде без перестановочных матриц, о привлечении способов факторизации, отличных от SVD.

## Литература

1. *Chen P.* Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix // *Int. J. Comput. Vision*, 2008. Vol. 80. Iss. 1. P. 125–142. doi: 10.1007/s11263-008-0135-7.
2. *Кривенко М. П.* Выбор модели при факторизации матрицы данных с пропусками // *Информатика и её применения*, 2022. Т. 16. Вып. 3. С. 52–58. doi: 10.14357/19922264220307.
3. *Abadir K. M., Magnus J. R.* Matrix algebra. — Cambridge: Cambridge University Press, 2005. 466 p.

4. *Hung C.-H., Markham T. L.* The Moore–Penrose inverse of a partitioned matrix  $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{B} & \mathbf{C} \end{pmatrix}$  // *Linear Algebra Appl.*, 1975. Vol. 11. Iss. 1. P. 73–86. doi: 10.1016/0024-3795(75)90118-4.
5. *Graybill F. A.* Matrices with applications in statistics. — Belmont, CA, USA: Wadsworth Publishing Co., 1983. 461 p.
6. *Seber G. A. F.* A matrix handbook for statisticians. — Hoboken, NJ, USA: Wiley, 2008. 569 p.

*Поступила в редакцию 01.12.22*

---

---

## EFFICIENT COMPUTATIONS IN MATRIX FACTORIZATION WITH MISSING COMPONENTS

*M. P. Krivenko*

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

**Abstract:** The paper is devoted to the effective implementation of matrix factorization in the presence of missing components into a product of two lower rank matrices. The problem of estimating the parameters of the adopted data model is solved by multidimensional optimization. In practice, the large sizes of the matrices and vectors included in iterative algorithms give rise to the curse of dimensionality. It is proposed to drastically reduce the complexity of matrix operations by presenting them in block-diagonal form. The article substantiates the possibility of casting individual matrices to a block-diagonal form and describes the rules for block-by-block singular value decomposition of matrices. The results of block-by-block processing are illustrated by the example of data matrix factorization of different sizes and with different probabilities of missing components. The time for estimating parameters can be reduced by several orders of magnitude compared to the processing of matrices in the usual representation.

**Keywords:** lower rank matrix approximation; singular decomposition; missing data; ALS algorithm; block-diagonal representation of a matrix

**DOI:** 10.14357/08696527230108

### References

1. Chen, P. 2008. Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix. *Int. J. Comput. Vision* 80(1):125–142. doi: 10.1007/s11263-008-0135-7.



2. Krivenko, M. P. Vybór modeli pri faktorizatsii matritsy dannykh s propuskami [Model selection for matrix factorization with missing components]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3):52–58. doi: 10.14357/19922264220307.
3. Abadir, K. M., and J. R. Magnus. 2005. *Matrix algebra*. Cambridge: Cambridge University Press. 466 p.
4. Hung, C.-H., and T. L. Markham. 1975. The Moore–Penrose inverse of a partitioned matrix  $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{B} & \mathbf{C} \end{pmatrix}$ . *Linear Algebra Appl.* 11(1):73–86. doi: 10.1016/0024-3795(75)90118-4.
5. Graybill, F. A. 1983. *Matrices with applications in statistics*. Belmont, CA: Wadsworth Publishing Co. 461 p.
6. Seber, G. A. F. 2008. *A matrix handbook for statisticians*. Hoboken, NJ: Wiley. 569 p.

*Received December 1, 2022*

### **Contributor**

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

## ОБ ОЦЕНКАХ УСТОЙЧИВОСТИ И ИХ ПРИМЕНЕНИИ ДЛЯ НЕКОТОРЫХ МОДЕЛЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ

*И. А. Ковалёв*<sup>1</sup>

**Аннотация:** Рассматриваются модели обслуживания, описываемые марковскими цепями с непрерывным временем. В статье применен один из методов, используемых для изучения устойчивости и получения соответствующих количественных оценок возмущений (неоднородных) цепей Маркова с непрерывным временем и конечным или счетным пространством состояний. Рассматриваются несколько конкретных моделей. Получены оценки устойчивости различных характеристик таких систем. Рассмотрены оценки, которые могут быть полезны для решения задач, связанных с управлением, а именно: с изменением мощности потока требований или мощности сервера, для того чтобы среднее число требований в системе находилось в заданных границах. Рассмотрен численный пример.

**Ключевые слова:** нестационарные системы обслуживания; марковские модели; оценки устойчивости; системы массового обслуживания; мощность потока; мощность сервера

**DOI:** 10.14357/08696527230109

### 1 Введение

В [1] впервые рассмотрены вопросы, связанные с устойчивостью неоднородных марковских цепей с непрерывным временем, более подробно для нестационарных процессов рождения и гибели в работах [2, 3]. В настоящей статье получены оценки устойчивости для двух систем.

Нестационарная марковская система с групповым поступлением и групповым обслуживанием требований с управлением, зависящим от состояния, изучалась в статьях [4–9]. Здесь рассмотрим частный случай из разд. 5 статьи [9].

Вторая модель — с одним сервером, специальным групповым поступлением и со специальной политикой пропуска очереди — впервые исследована в [10], соответствующая нестационарная ситуация была впервые изучена в [11]. Также модель (и ее различные вариации) обсуждалась в [9, 12]. В настоящей работе исследуются оценки устойчивости, когда интенсивности поступления групп требований экспоненциально убывают при увеличении размера группы. Также для нее рассмотрены оценки, которые могут быть полезны для решения задач, связанных с управлением, а именно: с изменением мощности потока требований или мощности сервера, для того чтобы среднее число требований в системе находилось в заданных границах.

---

<sup>1</sup>Вологодский государственный университет; Московский центр фундаментальной и прикладной математики, kovalev.iv96@yandex.ru

Применим метод получения оценок устойчивости нестационарных марковских цепей с непрерывным временем с использованием специальных весовых норм, связанных с полной вариацией, который описан в [13]. Метод основан на оценках, получаемых при помощи логарифмической нормы оператора линейного дифференциального уравнения, и специальных преобразованиях редуцированной матрицы интенсивностей процесса.

Рассмотрен численный пример построения предельных характеристик возмущенной системы обслуживания.

## 2 Основные понятия

Пусть  $X(t)$ ,  $t \geq 0$ , — нестационарная, вообще говоря, марковская цепь с непрерывным временем и счетным пространством состояний  $\{0, 1, 2, \dots\}$ . Обозначим через

$$p_{ij}(s, t) = \mathbf{P}\{X(t) = j | X(s) = i\},$$

где  $i, j \geq 0$ ,  $0 \leq s \leq t$ , переходные вероятности для  $X = X(t)$ . Пусть  $p_i(t) = \mathbf{P}\{X(t) = i\}$  — вероятности состояний цепи, а  $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$  — вектор распределения вероятностей состояний в момент времени  $t$ . Далее предполагается, что для любых  $i, j$

$$\mathbf{P}(X(t+h) = j | X(t) = i) = \begin{cases} q_{ij}(t)h + \alpha_{ij}(t, h), & \text{если } i \neq j; \\ 1 - \sum_{i \neq j} q_{ij} + \alpha_i(t, h), & \text{если } i = j, \end{cases}$$

где  $\sup_i |\alpha_i(t, h)| = o(h)$ .

Все интенсивности предполагаются локально интегрируемыми на промежутке  $[0, \infty)$ , а также  $a_{ij}(t) = q_{ji}(t)$  для  $i \neq j$  и  $a_{ii}(t) = -\sum_{i \neq j} a_{ji}(t) = -\sum_{i \neq j} q_{ij}(t)$ .

Также предполагаем выполненным условие  $|a_{ii}| \leq L < \infty$  при всех  $i$  и почти всех  $t \geq 0$ . Тогда получаем для описания процесса прямую систему Колмогорова

$$\frac{d\mathbf{p}(t)}{dt} = A(t)\mathbf{p}(t), \quad (1)$$

где  $A(t) = Q^T(t)$  — транспонированная матрица интенсивностей. Далее через  $\|\cdot\|$  обозначим  $l_1$ -норму, а именно:  $\|\mathbf{x}\| = \sum |x_i|$  и  $\|B\| = \sup_j \sum_i |b_{ij}|$  для матрицы  $B = (b_{ij})_{i,j=-\infty}^{\infty}$ . Рассмотрим  $\Omega$  — множество всех стохастических векторов, т. е.  $l_1$ -векторов с неотрицательными координатами и единичной нормой. Следовательно,

$$\|A(t)\| \leq 2 \sup_k |a_{kk}(t)| \leq 2L$$

при почти всех  $t \geq 0$ . Тогда оператор  $A(t)$  из пространства  $l_1$  в себя ограничен при почти всех  $t \geq 0$  и локально интегрируем на  $[0, \infty)$ . Таким образом, можно

рассмотреть (1) как дифференциальное уравнение в пространстве последовательностей  $l_1$  с ограниченным оператором.

Можно положить  $p_0(t) = 1 - \sum_{i \geq 1} p_i(t)$ . Тогда из (1) получаем следующее уравнение:

$$\frac{dz(t)}{dt} = B(t)z(t) + f(t),$$

где

$$B(t) = (b_{ij}(t))_{i,j}^\infty = (a_{ij}(t) - a_{i0}(t))_{i,j}^\infty; \quad f(t) = (a_{10}, a_{20}, \dots)^T.$$

Через  $E(t, k) = E\{X(t)|X(0) = k\}$  будем обозначать далее среднее (математическое ожидание) процесса в момент  $t$  при начальном условии  $X(0) = k$ .

### 3 Теорема об устойчивости

Пусть  $\{d_i\}$ ,  $i = 1, 2, \dots$ , — возрастающая последовательность положительных чисел, причем  $d_1 = 1$ . Рассмотрим треугольную матрицу  $D$  вида

$$\begin{pmatrix} d_1 & d_1 & d_1 & \dots \\ 0 & d_2 & d_2 & \dots \\ 0 & 0 & d_3 & \dots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

и соответствующее пространство последовательностей  $l_{1D} = \{z(t) = (p_1(t), p_2(t), \dots)^T / \|z\|_{1D} = \|z\| < \infty\}$ . Отметим, что  $\|B\|_{1D} = \|DBD^{-1}\|$ .

**Определение 1.** Марковская цепь  $X(t)$  называется  $1D$ -экспоненциально слабо эргодичной, если существуют положительные  $M$  и  $a$  такие, что при любых  $s$  и  $t$  таких, что  $t \geq s \geq 0$ , и любых начальных условиях  $\mathbf{p}^*(s) \in l_{1D}$ ,  $\mathbf{p}^{**}(s) \in l_{1D}$  выполнено неравенство:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1D} \leq M e^{-a(t-s)} \|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\|_{1D}.$$

**Определение 2.** Марковская цепь  $X(t)$  имеет предельное математическое ожидание (предельное среднее)  $\phi(t)$ , если  $|E(t, k) - \phi(t)| \rightarrow 0$  при  $t \rightarrow \infty$  для любого  $k$ .

Предполагаем, что выполнены неравенства:

$$\left. \begin{aligned} \|B(t)\|_{1D} \leq \mathfrak{B} \leq \infty; & \quad \|f(t)\|_{1D} \leq \mathfrak{f} \leq \infty; \\ \|B(t) - \bar{B}(t)\|_{1D} \leq |\mathfrak{B} - \bar{\mathfrak{B}}|; & \quad \|f(t) - \bar{f}(t)\|_{1D} \leq |\mathfrak{f} - \bar{\mathfrak{f}}| \end{aligned} \right\} \quad (2)$$

для всех  $t \geq 0$ .

**Теорема 1.** Пусть марковская цепь  $X(t)$   $1D$ -экспоненциально слабо эргодична. Тогда при достаточно малых возмущениях (2) возмущенная цепь  $\bar{X}(t)$

также  $1D$ -экспоненциально слабо эргодична и справедлива следующая оценка устойчивости в  $1D$ -норме:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\|_{1D} \leq \frac{M(M|\mathfrak{B} - \bar{\mathfrak{B}}|f + a|f - \bar{f}|)}{a(a - M|\mathfrak{B} - \bar{\mathfrak{B}}|)}.$$

Более того, если  $W = \inf_{k \geq 1} (d_k/k) > 0$ , то обе цепи Маркова  $X(t)$  и  $\bar{X}(t)$  имеют предельные математические ожидания и

$$\limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| \leq \frac{M(M|\mathfrak{B} - \bar{\mathfrak{B}}|f + a|f - \bar{f}|)}{Wa(a - M|\mathfrak{B} - \bar{\mathfrak{B}}|)}.$$

**Следствие 1.** Пусть  $X(t)$  — нестационарная  $1D$ -экспоненциально слабо эргодичная марковская цепь. Тогда справедлива следующая оценка устойчивости:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| \leq \frac{4M(M|\mathfrak{B} - \bar{\mathfrak{B}}|f + a|f - \bar{f}|)}{a(a - M|\mathfrak{B} - \bar{\mathfrak{B}}|)}.$$

Доказательства можно посмотреть в [13].

#### 4 Нестационарная марковская система с групповым поступлением и групповым обслуживанием требований с управлением, зависящим от состояния

В статье [8] рассмотрена марковская система массового поступления и массового обслуживания с общим управлением, зависящим от состояния [4–8]. Число требований в момент времени  $t$  в этой системе  $X(t)$  составляет цепь Маркова с непрерывным временем и пространством состояний  $\{0, 1, 2, \dots\}$ . Матрица интенсивностей  $Q(t) = (q_{ij}(t))_{i,j=0}^{\infty}$  имеет довольно специфическую структуру, а именно:

$$q_{ij}(t) = \begin{cases} h_{ij}(t), & \text{если } 0 \leq i \leq k-1, j \geq 0; \\ b_{i-j+k}(t), & \text{если } i \geq k, j \geq i-k; \\ 0 & \text{иначе,} \end{cases}$$

где

$$\begin{aligned} h_{ij}(t) &\geq 0 \quad (i \neq j); \\ 0 \leq -h_{ii}(t) &= \sum_{j \neq i} h_{ij}(t) < +\infty \quad (0 \leq i \leq k-1); \\ b_0(t) &> 0, \quad b_j(t) \leq 0 \quad (j \neq k); \\ \sum_{j=N+1}^{\infty} b_j(t) &> 0; \quad 0 \leq -b_k(t) = \sum_{j \neq k} b_j(t) < +\infty \end{aligned}$$

для некоторого фиксированного целого  $k \geq 1$ .

В данной статье получим оценки устойчивости для частного случая этой системы. В частности, возьмем модель из разд. 5 статьи [9], где была получена оценка скорости сходимости. Элементы матрицы интенсивностей  $Q(t)$  следующие:

$$h_{i,i-1}(t) = \mu(t); \quad h_{i,i+1}(t) = \lambda(t); \quad h_{i,i}(t) = -(\lambda(t) + \mu(t)); \\ b_0 = \mu; \quad b_{k+1} = \lambda(t); \quad b_k = -(\lambda(t) + \mu(t)); \quad k = 3,$$

где  $\lambda(t)$  и  $\mu(t)$  неотрицательны и локально интегрируемы по  $[0, \infty)$ . Тогда транспонированная матрица интенсивностей  $A(t) = (a_{ij}(t))_{i,j=0}^{\infty} = Q^T(t)$  имеет вид:

$$A(t) = \begin{pmatrix} -\lambda(t) & \mu(t) & 0 & \mu(t) \\ \lambda(t) & -(\lambda(t) + \mu(t)) & \mu(t) & 0 \\ 0 & \lambda(t) & -(\lambda(t) + \mu(t)) & 0 \\ 0 & 0 & \lambda(t) & -(\lambda(t) + \mu(t)) \\ 0 & 0 & 0 & \lambda(t) \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 & \dots \\ & & \mu(t) & 0 & \dots \\ & & 0 & \mu(t) & \dots \\ & & 0 & 0 & \dots \\ & & -(\lambda(t) + \mu(t)) & 0 & \dots \\ & & \lambda(t) & -(\lambda(t) + \mu(t)) & \dots \\ & & \vdots & \vdots & \ddots \end{pmatrix} \quad (3)$$

Возьмем  $\sigma > 1$  и возрастающую последовательность положительных чисел  $\{d_n, n \geq 1\}$  как  $d_n = \sigma^{n-1}$ . Отметим, что ранее в статье [9] получено следующее утверждение.

**Утверждение 1.** Если  $\int_0^{\infty} (\mu(t)(1 - \sigma^{-1}) - \lambda(t)(\sigma - 1)) dt = +\infty$  для некоторого  $\sigma > 1$ , тогда цепь Маркова  $X(t)$  слабо эргодична и выполняется оценка

$$|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)| \leq 4e^{-\int_0^t (\mu(u)(1 - \sigma^{-1}) - \lambda(u)(\sigma - 1)) du} |\mathbf{p}^*(0) - \mathbf{p}^{**}(0)|_{1D}.$$

Рассмотрим процесс  $\bar{X}(t), t \geq 0$ , описывающий число требований в «возмущенной» системе обслуживания, с соответствующей транспонированной матрицей интенсивностей  $\bar{A}(t)$ , где матрица «возмущения»  $\hat{A}(t) = A(t) - \bar{A}(t)$  в некотором смысле мала и имеет ту же структуру с соответствующими возмущенными интенсивностями  $\bar{\lambda}(t)$  и  $\bar{\mu}(t)$ .

Пусть

$$|\lambda(t) - \bar{\lambda}(t)| = |\hat{\lambda}(t)| \leq \hat{\epsilon}; \quad |\mu(t) - \bar{\mu}(t)| = |\hat{\mu}(t)| \leq \hat{\epsilon}.$$

Тогда из (3) получаем следующее ограничение:

$$\|\hat{A}(t)\| = 2 \sup_k |\hat{a}_{kk}(t)| \leq 4\hat{\epsilon}$$

и из теоремы 1 следующие оценки.

**Теорема 2.** *Если процесс  $X(t)$   $1D$ -экспоненциально слабо эргодичен, т. е.*

$$e^{-\int_0^\infty (\mu(u)(1-\sigma^{-1}) - \lambda(u)(\sigma-1)) du} \leq M e^{-a(t-s)}$$

для некоторых положительных  $M$  и  $a$ , то  $\bar{X}(t)$  также  $1D$ -экспоненциально слабо эргодичен. Пусть

$$H = \sup_{|i-j|=1} \frac{d_i}{d_j} < \infty.$$

Тогда выполняется следующая оценка устойчивости:

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\|_{1D} \leq \frac{M\hat{\epsilon}(2M(H+1)L+a)}{a(a-2M\hat{\epsilon}(H+1))}.$$

Более того, если  $W = \inf_{k \geq 1} (d_{k+1}/k) > 0$ , то

$$\limsup_{t \rightarrow \infty} |E(t, 0) - \bar{E}(t, 0)| \leq \frac{M\hat{\epsilon}(2M(H+1)L+a)}{Wa(a-2M\hat{\epsilon}(H+1))}.$$

## 5 Модель с одним сервером, специальным групповым поступлением и со специальной политикой пропуска очереди

Первоначальное описание исходной модели и первые исследования проведены в [10], соответствующая нестационарная ситуация была впервые изучена в [11]. Также модель (и ее различные вариации) обсуждалась в [9, 12]. При этом число требований в соответствующей системе обслуживания описывается процессом  $X(t)$  (см. далее). В указанных работах рассматриваются оценки эргодичности. В настоящей работе исследуются оценки устойчивости, когда интенсивности поступления групп требований экспоненциально убывают при увеличении размера группы.

Отметим, что рассматриваемый процесс  $X(t)$  является неоднородной марковской цепью с непрерывным временем, счетным пространством состояний  $\{0, 1, 2, \dots\}$ , а транспонированная матрица интенсивностей  $A(t) = (a_{ij}(t))_{i,j=0}^\infty$  имеет вид:

$$A(t) = \begin{pmatrix} -\lambda(t) & \mu(t) & 0 & 0 & \dots \\ \lambda(t)b_1 - (\lambda(t)B_2 + \mu(t)) & \mu(t) & 0 & 0 & \dots \\ \lambda(t)b_2 & \lambda(t)b_2 & -(\lambda(t)B_3 + \mu(t)) & \mu(t) & \ddots \\ \lambda(t)b_3 & \lambda(t)b_3 & \lambda(t)b_3 & -(\lambda(t)B_4 + \mu(t)) & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Здесь  $\lambda(t)$  — интенсивность поступления группы требований;  $\mu(t)$  — интенсивность обслуживания одного требования;  $\gamma(t)$  — интенсивность катастрофы (одномоментной потери всех требований в системе). Далее,  $\lambda(t)b_k$  — интенсивность поступления группы требований такая, что общее их число в системе оказывается равным  $k$ , т. е. если в системе уже есть  $k - j$  требований, то это интенсивность одновременного поступления группы из  $j$  требований. При этом  $B_k = \sum_{j=k}^{\infty} b_j$  при всех  $k \geq 1$  и  $B_1 = 1$ . Все функции, определяющие интенсивности, предполагаются неотрицательными и локально интегрируемыми на  $[0, \infty)$ . Кроме того, предполагается, что существуют  $C > 0$  и  $0 \leq q \leq 1$  такие, что

$$b_k < Cq^k.$$

Рассматриваем оценки устойчивости основных характеристик процесса  $X(t)$  при таких возмущениях, предполагая дополнительно, что процесс  $X(t)$  экспоненциально эргодичен, т. е. что для некоторых положительных  $c, b$ , последовательности  $\delta < 1, d_k = \delta^{-k}$  и всех  $s, t, 0 \leq s \leq t$ , выполняется неравенство:

$$e^{-\int_0^t (1-\delta)\mu(u) du} \leq Me^{-a(t-s)}. \tag{4}$$

Пусть возмущенные интенсивности  $\bar{\lambda}(t)$  и  $\bar{\mu}(t)$  такие, что  $|\lambda(t) - \bar{\lambda}(t)| \leq \hat{\epsilon}_\lambda$  и  $|\mu(t) - \bar{\mu}(t)| \leq \hat{\epsilon}_\mu$  для всех  $t \geq 0$ . Тогда

$$\|B(t)\|_{1D} \leq \|\lambda(t)\| + (1 + \delta)\|\mu(t)\| \leq (2 + \delta)L;$$

$$\|f(t)\|_{1D} \leq C^*L;$$

$$\|B(t) - \bar{B}(t)\|_{1D} \leq |\lambda(t) - \bar{\lambda}(t)| + (1 + \delta)\|\mu(t) - \bar{\mu}(t)\| \leq \hat{\epsilon}_\lambda + (1 + \delta)\hat{\epsilon}_\mu;$$

$$\|f(t) - \bar{f}(t)\|_{1D} \leq C^*\hat{\epsilon}_\lambda$$

для всех  $t \geq 0$ .

**Теорема 3.** Пусть  $X(t)$  —  $1D$ -экспоненциально слабо эргодическая цепь Маркова, т. е. выполняется неравенство (4). Тогда возмущенный процесс  $\bar{X}(t)$  также является  $1D$ -экспоненциально слабо эргодичным для достаточно малых  $\hat{\epsilon}_\lambda$  и  $\hat{\epsilon}_\mu$  и выполняется следующая оценка возмущения:



$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\|_{1D} \leq \frac{MC^*(ML((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda) + a\hat{\epsilon}_\lambda)}{a(a - M((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda))},$$

где  $M$  и  $a$  определяются (4). Более того, если  $W = \inf_{k \geq 1} (d_k/k) > 0$ , то обе цепи Маркова  $X(t)$  и  $\bar{X}(t)$  имеют предельные математические ожидания и

$$\limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| \leq \frac{MC^*(ML((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda) + a\hat{\epsilon}_\lambda)}{Wa(a - M((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda))}.$$

Используя хорошо известные соотношения между нормами:

$$\|p^* - p^{**}\| \leq 2\|z^* - z^{**}\| \leq 4\|z^* - z^{**}\|_{1D},$$

где  $z^*$  и  $z^{**}$  являются соответствующими векторами для  $p^*$  и  $p^{**}$ , получаем оценку возмущения в «естественной»  $l_1$ -норме (всего вариация, другими словами).

**Следствие 1.** Пусть  $X(t)$  — неоднородная  $1D$ -экспоненциально слабо эргодичная цепь Маркова. Тогда выполняется следующая оценка возмущения:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| \leq \frac{4MC^*(ML((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda) + a\hat{\epsilon}_\lambda)}{a(a - M((1 + \delta)\hat{\epsilon}_\mu + \hat{\epsilon}_\lambda))}.$$

**Замечание.** При исследовании этой модели на устойчивость по отношению к возмущениям  $b_k$  нужны дополнительные требования на малость возмущений во взвешенной норме. В то же время просто малые изменения  $b_k$  могут все сломать. Например, если взять  $\lambda = \mu = 1$ , а  $b_1 = 1 - \varepsilon$ ,

$$b_k = \begin{cases} 0 & \text{при } 1 < k \leq N; \\ \frac{\varepsilon}{(k - N)(k - N + 1)} & \text{при } k > N, \end{cases}$$

то при возрастании  $N$  норма возмущения в  $l_1$  будет  $\varepsilon$ . Однако при этом уже не выполнено условие  $\sum_k B_k < \infty$ .

## 6 Получение границ для некоторых параметров модели из раздела 5

В данном разделе рассмотрим задачи на управление.

**Задача 1.** Считая, что известны среднее число требований в системе  $E(t, 0)$  с границами и интенсивность поступления требований  $\lambda(t)$ , получить границы для интенсивности обслуживания требований  $\mu(t)$ . Решив эту задачу, можно понять, насколько надо увеличить интенсивность обслуживания, чтобы среднее уменьшилось на  $v\%$  (или, наоборот, насколько можно уменьшить интенсивность обслуживания, чтобы среднее не выросло больше чем на  $v\%$ ).

Для ее решения полезны оценки устойчивости, а именно: теорема 3. Сформулируем из нее утверждение.

**Утверждение 2.** Пусть задано среднее число требований в системе  $E(t, 0)$  с некоторыми границами, т. е.

$$\limsup_{t \rightarrow \infty} |E(t, 0) - \bar{E}(t, 0)| \leq h$$

для некоторого положительного  $h$ . Тогда границы для интенсивности обслуживания будут иметь вид:

$$|\mu(t) - \bar{\mu}(t)| \leq \frac{Wa^2h}{(1 + \delta)(C^*LM^2 + MhWa)}.$$

**Задача 2.** Считая, что известны среднее число требований в системе  $E(t, 0)$  с границами и интенсивность обслуживания требований  $\mu(t)$ , получить границы для интенсивности поступления требований  $\lambda(t)$ . Решив эту задачу, можно понять, насколько надо уменьшить интенсивность поступления, чтобы среднее уменьшилось на  $v\%$  (или, наоборот, насколько можно увеличить интенсивность поступления, чтобы среднее не выросло больше чем на  $v\%$ ).

**Утверждение 3.** Пусть задано среднее число требований в системе  $E(t, 0)$  с некоторыми границами, т. е.

$$\limsup_{t \rightarrow \infty} |E(t, 0) - \bar{E}(t, 0)| \leq h$$

для некоторого положительного  $h$ . Тогда границы интенсивности поступления требований можно оценить следующим образом:

$$|\lambda(t) - \bar{\lambda}(t)| \leq \frac{Wa^2h}{MC^*(ML + a) + WahM}.$$

**Задача 3.** Определить мощность сервера, которая гарантирует, что в долгосрочной перспективе среднее число требований в системе не будет превышать  $N^*$ .

Будем считать, что интенсивность обслуживания задается выражением  $\mu(t) = \mu g(t)$ , где  $g(t)$  известно,  $e^{-\int_{\tau}^t g(u)du} \leq He^{-v(t-\tau)}$ , а  $\mu$  (мощностью сервера) можем управлять. Оценим параметры, связанные с обслуживанием, а именно: мощность  $\mu$  сервера, считая, что остальные параметры заданы и среднее не должно превосходить  $N^*$ , т. е.  $\limsup_{t \rightarrow \infty} E(t, 0) \leq N^*$ .

По определению

$$E(t, 0) = \sum_{n \in \mathcal{X}} np_n(t).$$

Поскольку все  $p_i(t)$  неотрицательны, имеем

$$\|\mathbf{z}\|_{1D} = \sum_{n \geq 1} p_n \sum_{k \geq 1} d_k \geq \sum_{n \geq 1} d_n p_n \geq W \sum_{n \geq 1} n p_n, \quad (5)$$

где  $W = \inf_{k \geq 1} (d_k/k) > 0$ . Тогда  $E(t, 0) \leq 1/(W)\|\mathbf{z}(t)\|_{1D}$ .

С другой стороны, получаем следующее неравенство:

$$\begin{aligned} \|\mathbf{z}\|_{1D} &\leq \|V(t)\|_{1D} \|\mathbf{z}(0)\|_{1D} + \int_0^t \|V(t, \tau)\|_{1D} \|\mathbf{f}(\tau)\|_{1D} d\tau \leq \\ &\leq 4e^{-(1-\delta)\mu \int_0^t g(u) du} \|\mathbf{z}(0)\|_{1D} + 4C^* L \int_0^t e^{-(1-\delta)\mu \int_\tau^t g(u) du} d\tau \leq \\ &\leq 4(H e^{-vt})^{(1-\delta)\mu} \|\mathbf{z}(0)\|_{1D} + 4C^* L \int_0^t \left( H e^{-v(t-\tau)} \right)^{(1-\delta)\mu} d\tau \leq \frac{4C^* L H^{(1-\delta)\mu}}{v(1-\delta)\mu}. \end{aligned}$$

Тогда

$$\limsup_{t \rightarrow \infty} E(t, 0) \leq \frac{4C^* L H^{(1-\delta)\mu}}{W v(1-\delta)\mu} \leq N^*$$

и получаем следующее утверждение.

**Теорема 4.** Пусть задана интенсивность поступления требований  $\lambda(t)$ , известны  $\{b_k\}$ . Интенсивность обслуживания имеет вид  $\mu(t) = \mu g(t)$ , причем  $g(t)$  известна и выполняется неравенство  $e^{-\int_\tau^t g(u) du} \leq H e^{-v(t-\tau)}$ , а  $\mu$  можем управлять. Тогда для того, чтобы

$$\limsup_{t \rightarrow \infty} E(t, 0) \leq N^*,$$

достаточно, чтобы

$$\frac{H^{(1-\delta)\mu}}{\mu} \leq \frac{v(1-\delta)N^*W}{4C^*L}.$$

**Задача 4.** Определить мощность потока требований, которая гарантирует, что в долгосрочной перспективе среднее число требований в системе не будет превышать  $N^*$ .

Будем считать, что интенсивность поступления задается выражением  $\lambda(t) = \lambda g(t)$ , в котором  $g(t)$  известна и  $|g(t)| \leq H$ , а  $\lambda$  (мощность потока поступающих требований) можем управлять. Оценим параметры, связанные с поступлением,

а именно: мощность потока  $\lambda$ , считая, что остальные параметры заданы и среднее не должно превосходить  $N^*$ , т. е.  $\limsup_{t \rightarrow \infty} E(t, 0) \leq N^*$ .

С одной стороны, имеем (5), а с другой стороны, верно следующее неравенство:

$$\begin{aligned} \|z\|_{1D} &\leq \|V(t)\|_{1D} \|z(0)\|_{1D} + \int_0^t \|V(t, \tau)\|_{1D} \|f(\tau)\|_{1D} d\tau \leq \\ &\leq 4e^{-(1-\delta)Lt} \|z(0)\|_{1D} + 4C^* H \lambda \int_0^t e^{-(1-\delta)L(t-\tau)} d\tau \leq \frac{4C^* H \lambda}{(1-\delta)L}. \end{aligned}$$

Следовательно,

$$E(t, 0) \leq \frac{4C^* H \lambda}{(1-\delta)LW} \leq N^*$$

и

$$\lambda \leq \frac{(1-\delta)LN^*W}{4C^*H}.$$

**Теорема 5.** Пусть задана интенсивность обслуживания требований  $\mu(t)$ . Интенсивность поступления имеет вид  $\lambda(t) = \lambda g(t)$ , причем  $g(t)$  известна и  $|g(t)| \leq H$ . Тогда для того, чтобы

$$\limsup_{t \rightarrow \infty} E(t, 0) \leq N^*,$$

достаточно, чтобы

$$\lambda \leq \frac{(1-\delta)LN^*W}{4C^*H}.$$

## 7 Пример

В примере, представленном ниже, предполагается, что распределение групп по размеру  $\{b_k, k \geq 1\}$  — геометрическое, т. е.  $b_k = (1-q)q^{k-1}$ ,  $k \geq 1$ ,  $0 < q < 1$ . Также предполагается, что интенсивности поступления и/или обслуживания — периодические функции. Учитывая, что  $\{b_k, k \geq 1\}$  — геометрическое, независимо от параметра геометрического распределения периодические интенсивности гарантируют существование (периодического) предельного распределения  $X(t)$ .

Рассмотрим пример 1 из [11]. Для него уже получены оценки скорости сходимости, поэтому на них подробно останавливаться не будем.

Пусть очередь имеет следующие интенсивности перехода:

$$\lambda(t) = 1 + \sin(2\pi t); \quad \mu(t) = 1 + \cos(2\pi t).$$

Пусть  $q = 2/3$ , т. е. распределение групп по размерам имеет вид:  $b_k = 2^{k-1}/3^k$ ,  $k \geq 1$ , тогда средний размер группы  $\sum_{k=1}^{\infty} k b_k = 3$ . Положим  $\delta = 5/6$ . Оценка скорости сходимости имеет вид:

$$e^{-\int_s^t \alpha^*(u) du} = e^{-(1/6) \int_s^t (1 + \cos(2\pi u)) du} \leq 2e^{-(1/6)(t-s)}.$$

Следовательно, можно взять  $a = 1/6$ ,  $M = 2$ . Очевидно, что  $L \leq 2$ ,  $C = 1/3$ , а  $C^* = 10/3$ . Также  $W = \inf_{k \geq 1} d_k/k = 3/4$ .

Из теоремы 3 получаем соответствующие оценки устойчивости:

$$\begin{aligned} \limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\|_{1D} &\leq \frac{40(44\hat{\epsilon}_\mu + 25\hat{\epsilon}_\lambda)}{1 - 22\hat{\epsilon}_\mu - 12\hat{\epsilon}_\lambda}; \\ \limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| &\leq \frac{160(44\hat{\epsilon}_\mu + 25\hat{\epsilon}_\lambda)}{3(1 - 22\hat{\epsilon}_\mu - 12\hat{\epsilon}_\lambda)}; \\ \limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| &\leq \frac{160(44\hat{\epsilon}_\mu + 25\hat{\epsilon}_\lambda)}{1 - 22\hat{\epsilon}_\mu - 12\hat{\epsilon}_\lambda}. \end{aligned}$$

Пусть  $|\mathbf{E}(t, 0) - \bar{\mathbf{E}}(t, 0)| \leq h = 5$ . Тогда из утверждения 3 границы для интенсивности поступления будут иметь вид:

$$|\lambda(t) - \bar{\lambda}(t)| \leq \hat{\epsilon}_\lambda \leq 10^{-2}.$$

Неравенство

$$E(t, 0) \leq \frac{4C^*H\lambda}{(1 - \delta)LW}$$

гарантирует, что среднее число требований в системе не превышает 107. Из теоремы 5 следует, что для того, чтобы среднее число требований в системе не превышало, например, 60, достаточно уменьшить мощность потока поступающих требований в 2 раза.

Неравенство

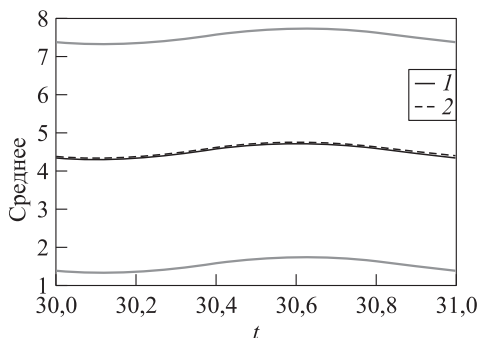
$$\mathbf{E}(t, 0) \leq \frac{4C^*LH^{(1-\delta)\mu}}{Wv(1 - \delta)\mu}$$

гарантирует, что среднее число требований в системе не превышает 1437. Из теоремы 4 следует, что для того, чтобы среднее число требований в системе не превышало, например, 500, достаточно увеличить мощность сервера в 5 раз.

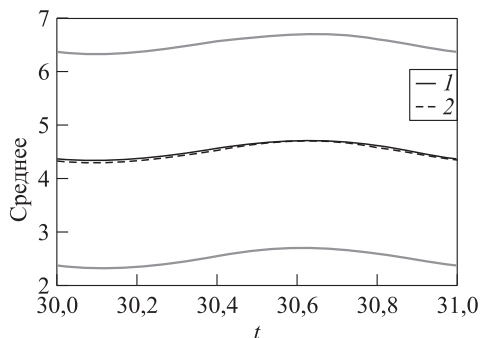
Пусть  $|\mathbf{E}(t, 0) - \bar{\mathbf{E}}(t, 0)| \leq h = 3$ . Тогда из утверждения 4 границы интенсивности обслуживания будут иметь вид:

$$|\mu(t) - \bar{\mu}(t)| \leq \hat{\epsilon}_\mu \leq \frac{3h}{7040 + 66h} \leq 10^{-2}.$$

На рис. 1 показано поведение среднего значения числа требований при  $\mu(t) = 0,99(1 + \cos(2\pi t))$ . На рис. 2 показано поведение среднего числа требований в системе, когда  $\lambda(t) = 0,99(1 + \sin(2\pi t))$ .



**Рис. 1** Среднее число требований  $E(t; 0)$  при  $\mu = 1$  (1) и 0,99 (2) для  $t \in [30, 31]$



**Рис. 2** Аппроксимация предельного среднего значения  $E(t, 0)$  с границами для  $\lambda = 1$  (1) и 0,99 (2) на  $t \in [30, 31]$

## Литература

1. Zeifman A. I. Stability for continuous-time nonhomogeneous Markov chains // Stability problems for stochastic models / Eds. V. V. Kalashnikov, V. M. Zolotarev. — Lecture notes in mathematics ser. — Berlin, Heidelberg: Springer, 1985. Vol. 1155. P. 401–414. doi: 10.1007/BFb0074830.
2. Zeifman A. Stability of birth and death processes // J. Mathematical Sciences, 1998. Vol. 91. P. 3023–3031. doi: 10.1007/BF02432876.
3. Андреев Д., Елесин М., Кузнецов А., Крылов Е., Зейфман А. Эргодичность и устойчивость нестационарных систем обслуживания // Теория вероятностей и математическая статистика, 2003. Т. 68. С. 1–11.
4. Chen A., Renshaw E. Markovian bulk-arriving queues with state-dependent control at idle time // Adv. Appl. Probab., 2004. Vol. 36. P. 499–524. doi: 10.1239/aap/1086957583.
5. Li J., Chen A. Decay property of stopped Markovian bulk-arriving queues // Adv. Appl. Probab., 2008. Vol. 40. P. 95–121.
6. Chen A., Pollett P., Li J., Zhang H. Markovian bulk-arrival and bulk-service queues with state-dependent control // Queueing Syst., 2010. Vol. 64. P. 267–304. doi: 10.1007/s11134-009-9162-5.
7. Chen A., Li J., Hou Z., Ng K. W. Decay properties and quasi-stationary distributions for stopped Markovian bulk-arrival and bulk-service queues // Queueing Syst., 2010. Vol. 66. P. 275–311. doi: 10.1007/s11134-010-9194-x.
8. Chen A., Wu X., Zhang J. Markovian bulk-arrival and bulk-service queues with general state-dependent control // Queueing Syst., 2020. Vol. 95 P. 331–378. doi: 10.1007/s11134-020-09660-0.
9. Zeifman A., Satin Y., Kovalev I., Razumchik R., Korolev V. Facilitating numerical solutions of inhomogeneous continuous time Markov chains using ergodicity bounds obtained with logarithmic norm method // Mathematics, 2021. Vol. 9. No. 1. Art. 42. doi: 10.3390/math9010042.

10. *Marin A., Rossi S.* A queueing model that works only on the biggest jobs // Computer performance engineering / Eds. M. Gribaudo, M. Iacono, T. Phung-Duc, R. Razumchik. — Lecture notes in computer science ser. — Cham: Springer, 2020. Vol. 12039. P. 118–132. doi: 10.1007/978-3-030-44411-2\_8.
11. *Zeifman A. I., Razumchik R. V., Satin Y. A., Kovalev I. A.* Ergodicity bounds for the Markovian queue with time-varying transition intensities, batch arrivals and one queue skipping policy // Appl. Math. Comput., 2021. Vol. 395. Art. 125846. 11 p. doi: 10.1016/j.amc.2020.125846.
12. *Зейфман А. И., Сатин Я. А., Ковалёв И. А.* Об одной нестационарной модели обслуживания с катастрофами и тяжелыми хвостами // Информатика и её применения, 2021. Т. 15. Вып. 2. С. 20–25. doi: 10.14357/19922264210203.
13. *Zeifman A., Korolev V., Satin Y.* Two approaches to the construction of perturbation bounds for continuous-time Markov chains // Mathematics, 2020. Vol. 8. No. 2. Art. 253. doi: 10.3390/math8020253.

Поступила в редакцию 13.02.23

---

---

## ON THE PERTURBATION BOUNDS AND THEIR APPLICATION FOR SOME QUEUEING MODELS

*I. A. Kovalev*<sup>1,2</sup>

<sup>1</sup>Department of Applied Mathematics, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

<sup>2</sup>Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

**Abstract:** The service models described by Markov chains with continuous time are considered. One of the known methods is used to study the perturbation and obtain appropriate quantitative of perturbations bounds of (inhomogeneous) Markov chains with continuous time and finite or countable state space. Several specific models are considered. The perturbation bounds of various characteristics of such systems are obtained. The bounds are also considered that can be useful for solving management-related tasks, namely, associated with changing the power of the flow of requirements or the server power so that the average number of requirements in the system is within the specified limits. A numerical example is considered.

**Keywords:** nonstationary service systems; Markov models; perturbation bounds; queuing systems; flow power; server power

**DOI:** 10.14357/08696527230109

### References

1. Zeifman, A. I. 1985. Stability for continuous-time nonhomogeneous Markov chains. *Stability problems for stochastic models*. Eds. V. V. Kalashnikov and V. M. Zolotarev. Lecture notes in mathematics ser. 1155:401–414. doi: 10.1007/BFb0074830.

2. Zeifman, A. 1998. Stability of birth and death processes. *J. Mathematical Sciences* 91:3023–3031. doi: 10.1007/BF02432876.
3. Andreev, D., M. Elesin, A. Kuznetsov, E. Krylov, and A. Zeyfman. 2003. Ergodichnost' i ustoychivost' nestatsionarnykh sistem obsluzhivaniya [Ergodicity and stability of nonstationary queueing system]. *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics] 68:1–11.
4. Chen, A., and E. Renshaw. 2004. Markovian bulk-arriving queues with state-dependent control at idle time. *Adv. Appl. Probab.* 36(2):499–524. doi: 10.1239/aap/1086957583.
5. Li, J., and A. Chen. 2008. Decay property of stopped Markovian bulk-arriving queues. *Adv. Appl. Probab.* 40(1):95–121.
6. Chen, A., P. Pollett, J. Li, and H. Zhang. 2010. Markovian bulk-arrival and bulk-service queues with state-dependent control. *Queueing Syst.* 64:267–304. doi: 10.1007/s11134-009-9162-5.
7. Chen, A., J. Li, Z. Hou, and K. W. Ng. 2010. Decay properties and quasi-stationary distributions for stopped Markovian bulk-arrival and bulk-service queues. *Queueing Syst.* 66:275–311. doi: 10.1007/s11134-010-9194-x.
8. Chen, A., X. Wu, and J. Zhang. 2020. Markovian bulk-arrival and bulk-service queues with general state-dependent control. *Queueing Syst.* 95:331–378. doi: 10.1007/s11134-020-09660-0.
9. Zeifman, A., Y. Satin, I. Kovalev, R. Razumchik, and V. Korolev. 2021. Facilitating numerical solutions of inhomogeneous continuous time Markov chains using ergodicity bounds obtained with logarithmic norm method. *Mathematics* 9(1):42. doi: 10.3390/math9010042.
10. Marin, A., and S. Rossi. 2020. A queueing model that works only on the biggest jobs. *Computer performance engineering*. Eds. M. Gribaudo, M. Iacono, T. Phung-Duc, and R. Razumchik. Lecture notes in computer science ser. Cham: Springer. 12039:118–132. doi: 10.1007/978-3-030-44411-2\_8.
11. Zeifman, A. I., R. V. Razumchik, Y. A. Satin, and I. A. Kovalev. 2021. Ergodicity bounds for the Markovian queue with time-varying transition intensities, batch arrivals and one queue skipping policy. *Appl. Math. Comput.* 395:125846. 11 p. doi: 10.1016/j.amc.2020.125846.
12. Zeifman, A. I., Y. A. Satin, and I. A. Kovalev. 2020. Ob odnoy nestatsionarnoy modeli obsluzhivaniya s katastrofami i tyazhelymi khvostami [On one nonstationary service model with catastrophes and heavy tails]. *Informatika i ee Primeneniya — Inform. Appl.* 15(2):20–25. doi: 10.14357/19922264210203.
13. Zeifman, A., I. Kovalev, and Y. Satin. 2020. Two approaches to the construction of perturbation bounds for continuous-time Markov chains. *Mathematics* 8(2):253. doi: 10.3390/math8020253.

*Received February 13, 2023*

## Contributor

**Kovalev Ivan A.** (b. 1996) — PhD student, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; kovalev.iv96@yandex.ru



## ПРОГРАММНОЕ ОПРЕДЕЛЕНИЕ СЕТИ В КОНВЕРГЕНТНОЙ И ГИПЕРКОНВЕРГЕНТНОЙ ИНФРАСТРУКТУРАХ

*В. Б. Егоров*<sup>1</sup>

**Аннотация:** В последнее время конвергентная (converged infrastructure, CI) и гиперконвергентная инфраструктуры (hyperconverged infrastructure, HCI) стали предметом широкого обсуждения, но сами эти термины не получили общепринятого определения. В статье делается попытка прояснить роль сети в такого рода инфраструктурах с учетом неопределенности с их трактовкой, а также оценить возможности и перспективы реализации в них сетей на принципах программного определения (software-defined networking, SDN). Основным принципам SDN соответствует такая существенная отличительная особенность CI, как централизованное управление всеми компонентами инфраструктуры с единой консоли. Однако это условие необходимо, но не достаточно, сеть типичной CI можно рассматривать лишь как шаг в направлении SDN. Полноценная реализация концепции SDN становится возможной в HCI, где из-за присущих гиперконвергенции особенностей поставщик фактически вынужден организовать в своей инфраструктуре программно-определяемое хранилище (software-defined storage, SDS) и потенциально обладает возможностью дополнить свой продукт SDN-сетью. Для рядовых владельцев центров обработки данных (ЦОД) покупка гиперконвергентной инфраструктуры может оказаться не только самым простым способом обретения современного SDS, но едва ли не единственно возможным — готовой SDN.

**Ключевые слова:** гиперконвергентная инфраструктура; конвергентная инфраструктура; программно-определяемые сети; CI; HCI; SDN

**DOI:** 10.14357/08696527230110

### 1 Введение

В области информационных технологий (ИТ) регулярно появляются новые понятия и модные термины. На волне моды они подхватываются и активно эксплуатируются рекламодателями, далеко не всегда заботящимися о точности терминологии, а порой и сущности рекламируемого объекта. Часто используемые термины и модные выражения не всегда получают четкое общепринятое определение, и их использование может вводить в заблуждение и приводить к взаимному непониманию. Подобного рода трудности были с программным определением сетей (SDN) [1, 2], похожие трудности возникают и с входящими ныне в моду конвергентным и гиперконвергентным подходами к инфраструктуре ЦОД. Новые модные термины конвергенции и гиперконвергенции в отношении

---

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vegorov@ipiran.ru

инфраструктуры ЦОД ИТ-специалисты и участники рынка стремятся понимать и трактовать так, как им выгодно [3].

Хотя конвергентная и гиперконвергентная инфраструктуры стали в последнее время объектами широкого обсуждения и активной рекламы, сами термины СІ и НСІ, будучи активно используемыми, тем не менее не получили общепринятого определения (*over-used and under-defined* [4]). Разные поставщики конвергентных инфраструктур используют для них различные названия: «конвергентная система» (*converged system*), «унифицированные вычисления» (*unified computing*), «динамическая инфраструктура» (*dynamic infrastructure*) и даже «центр обработки данных следующего поколения» [5]. Важный частный случай СІ также известен как «конвергентное хранилище». При этом под множеством разных названий на самом деле могут иметься в виду одни и те же или близкие по существу решения инфраструктуры ЦОД. И наоборот. Широкое использование модного термина НСІ в отсутствие четкого определения его сущности нередко вызывает путаницу даже у ИТ-специалистов.

Настоящая статья делает попытку прояснить роль и особенности сети в СІ и НСІ с учетом имеющейся неопределенности в трактовке этих терминов, а также оценить возможность реализации в такого рода инфраструктурах сетей на принципах SDN.

## 2 Сеть в конвергентной инфраструктуре

Словарь Oxford Languages определяет конвергенцию как «слияние каких-либо признаков и элементов в независимых друг от друга явлениях». Вероятно, исходно понятие конвергентности в отношении инфраструктуры ЦОД возникло в связи с идеей конвергентного хранилища и, как следствие, конвергентной сети в инфраструктуре ЦОД. Объединение в одном унифицированном аппаратном модуле прежде относительно независимых функциональных компонентов: вычислительных и хранилищных — предоставляет возможность проще компоновать и масштабировать инфраструктуру ЦОД, тиражируя такие унифицированные модули в качестве типовых «строительных блоков». Радикальное сокращение в СІ номенклатуры таких «блоков» дает возможность закупать все основное оборудование ЦОД у одного поставщика, производящего подходящие модули. В свою очередь, единственный поставщик обеспечивает конвергентному подходу еще два важных преимущества. Во-первых, у разработчика СІ как целостной системы появляется возможность сконцентрировать управление всем оборудованием на одной универсальной консоли, в том числе с использованием специфических фирменных средств и программных инструментов. Во-вторых, продавец получает возможность предварительно комплексно протестировать у себя поставляемое оборудование на совместимость и соответствие требованиям заказчика. В результате существенно сокращается время развертывания инфраструктуры на площадке покупателя и ускоряется ввод ЦОД в эксплуатацию, а для его обслуживания требуется меньше квалифицированного персонала. Вследствие

этих особенностей конвергентные инфраструктуры оказались особенно востребованными в малом и среднем бизнесе, а также в филиалах и удаленных офисах крупных компаний, где нередко имеются сложности с приобретением оборудования и обеспечением ЦОД ИТ-персоналом соответствующей квалификации.

Востребованность рождает предложение, и теперь свои готовые решения для СІ и НСІ предлагают на рынок многие фирмы, в числе которых как давно известные крупные игроки вроде Dell Technologies, Hewlett Packard Enterprise и Cisco Systems, так и молодые, специализирующиеся в этой области компании вроде Nutanix. Программному обеспечению (ПО) для СІ и НСІ уделяют серьезное внимание лидеры рынка ПО Microsoft и VMware.

При всем множестве трактовок и практических реализаций СІ в качестве наиболее общих и чаще всего присутствующих отличительных особенностей этой инфраструктуры можно выделить:

- организацию из типовых, поставляемых одним поставщиком модулей, которые включают вычислительные, хранилищные и сетевые компоненты;
- централизованное управление всеми компонентами с единой консоли с использованием фирменных программных инструментов.

В отсутствие четкого общепринятого определения СІ непросто оценить в ней роль сети как инфраструктурного компонента. Сложность дополнительно возрастает, если речь идет о сетях, организованных на принципах SDN, поскольку здесь неопределенность с терминологией и сущностью касается и инфраструктуры, и сети.

В типичных вариантах реализации СІ обязательно содержит вычислительные и хранилищные компоненты, а соответствующая аппаратура интегрируется в унифицированные модули инфраструктуры. Но сетевые компоненты при этом стоят особняком. В отличие от множественных и, что характерно для СІ, распределенных по всей инфраструктуре компонентов вычислений и хранения данных, сеть ЦОД, будучи их консолидирующим средством, обязана быть по своей сущности единой. Она не масштабируется горизонтально, и основная ее аппаратура (коммутаторы, маршрутизаторы и пр.) не интегрируется в типовые модули. В некоторых конвергентных решениях сеть вообще выносится за скобки инфраструктуры и по умолчанию считается неким внешним атрибутом с идеальными характеристиками [3, 6].

Между тем требования к сети в СІ могут оказаться существенно выше в сравнении с сетью традиционной инфраструктуры. Дело в том, что в типичных конвергентных системах в силу их организации может отсутствовать выделенное сетевое хранилище (network attached storage, NAS) или отдельная сеть хранения (storage-area network, SAN), а весь хранилищный трафик вынужденно переносится в общую конвергентную сеть ЦОД, в частности с использованием протоколов FibreChannel over Ethernet, RDMA over Converged Ethernet и пр. [7]. В результате, во-первых, существенно увеличивается суммарный трафик в сети ЦОД, а во-вторых, управление обоими компонентами конвергентного трафика, сетевым

и хранилищным, объективно требует тесной координации. Такая особенность СІ дополнительно стимулирует объединение управления всеми инфраструктурными компонентами на единой консоли.

Таким образом, управление сетью в СІ предполагается комплексно централизованным, что соответствует условию программной определяемости сети [2, 8]. Однако это только консольная централизация администрирования, но еще не оркестровка сети в процессе работы ЦОД. Хотя в некоторых публикациях СІ характеризуется как инфраструктура, «готовая к облаку» [9], вряд ли аналогичное утверждение справедливо в отношении ее готовности к SDN. Типичную сеть СІ более реалистично рассматривать лишь как шаг в направлении программного определения.

### 3 Сеть в гиперконвергентной инфраструктуре

В силу неопределенности с самими понятиями СІ и НСІ непросто и уловить разницу между ними. Ключевой здесь оказывается приставка «гипер», которая указывает на принципиальную для НСІ взаимосвязь конвергентной инфраструктуры с гипервизорами. Если СІ нередко позиционируется как аппаратно определяемая инфраструктура, то НСІ благодаря принципиальному вовлечению гипервизоров вправе претендовать на роль программно-определяемой.

Продвижение от СІ к НСІ возможно двумя путями: либо разработчик системного ПО каким-то образом приспособливает свой гипервизор к имеющимся на рынке СІ, либо продавец конвергентных систем разрабатывает собственный гипервизор и поставляет его вместе со своей инфраструктурой. Второй вариант в большей степени отвечает идее конвергенции с поставкой комплексного отработанного продукта единым поставщиком, целиком и полностью отвечающим за всю систему. Действительно, многие разработчики СІ идут по этому пути, благо сегодня на рынке ПО есть общедоступные гипервизоры с открытым кодом [5], которые необходимо лишь доработать под особенности конкретного оборудования и фирменную консоль.

Типичная НСІ компонуется из унифицированных модулей, в качестве которых используются серийно выпускаемые (off-the-shelf) серверы. В такой инфраструктуре концептуально нет места отдельному хранилищу данных типа NAS или SAN, оно физически реализуется распределенным множеством устройств дисковой и твердотельной памяти, подключенных непосредственно внутри модулей (direct attached storage, DAS). Это множество DAS программными средствами виртуализации объединяется в единый пул хранения данных, и для централизованного управления этим виртуальным пулом разработчиком НСІ создается специальное ПО, которое включает помимо консольной составляющей контроллер такого виртуализированного хранилища, работающий на гипервизорном уровне [4, 10]. В результате в НСІ фактически организуется программно-определяемое хранилище с фирменным контроллером, реализованным

как часть гипервизора. Именно SDS на основе множества DAS наиболее ярко отличает HCI от CI.

Концепция HCI в основном ориентирована на вычислительные и хранилищные компоненты, поскольку именно они интегрируются в унифицированном аппаратном модуле, а сетевым аспектам, как и в CI, нередко уделяется недостаточно внимания [11]. Между тем сеть в HCI требует особо тщательной проработки по тем же причинам, что и в CI: конвергентный сетевой трафик в HCI отличается повышенной интенсивностью в сравнении с трафиком в традиционных инфраструктурах. Поэтому для нормальной работы гиперконвергентной системы в ее сети требуются коммутаторы с высокоскоростными портами, минимальными задержками коммутации и приоритизацией трафика [7]. Задачу могли бы облегчить специальные коммутационные структуры, но такое решение входит в противоречие с концептуальной гомогенностью HCI [3].

Таким образом, сетевой компонент гиперконвергентной инфраструктуры оказывается в ней в значительной степени чужеродным и наиболее проблемным в реализации. Одной из заманчивых возможностей упростить задачу выглядит применение в HCI SDN-подхода [12]. Действительно, особенности HCI могут нивелировать некоторые слабые стороны последнего, препятствующие широкому внедрению SDN-сетей [1], а именно:

- желательность обзорности инфраструктуры сети для ее администратора, что в значительной мере ограничивает применимость SDN рамками ЦОД;
- необходимость оркестровки конкретной SDN-сети «по месту», что требует высококвалифицированных программистов и остается непосильным для большинства владельцев ЦОД.

Первое препятствие снимается в HCI самим ее предназначением, а более существенное второе может быть снято поставщиком инфраструктуры. Он, в отличие от рядового владельца ЦОД, обладает командой профессиональных программистов, которая уже встроила в его фирменный продукт централизованную консоль, разработала собственный гипервизор и организовала полноценный SDS. Наверняка этой команде окажется по плечу и дополнение фирменного продукта адекватной SDN-сетью, пусть даже не вполне соответствующей классическим критериям SDN. В частности, она, скорее всего, останется традиционной IP-сетью, лишь дополненной функциональностью протокола OpenFlow [13], а ее контроллер будет реализован не на прикладном уровне, а подобно контроллеру SDS внедрится в гипервизор. Но даже не будучи «классическим», SDN-подход к сети в HCI способен решать следующие задачи [11]:

- упрощать реконфигурирование и масштабирование сети;
- управлять физическими и виртуальными сетями в соответствии с заданным набором политик;
- облегчить переход в гибридное «облако».

Если CI представляет собой лишь шаг в направлении SDN, то HCI позволяет рядовым владельцам ЦОД ускоренно завершить этот путь и получить в каком-то готовом виде наряду с программно-определяемым хранилищем и программно-определяемую сеть. А имея SDS и SDN, можно ставить дальнейшую задачу создания целиком программно-определяемого ЦОД (software-defined data center, SDDC). Но это опять же задача не столько для рядовых владельцев ЦОД, сколько для разработчиков и поставщиков на рынок систем HCI, которые потенциально могли бы предложить покупателям готовые гиперконвергентные SDDC.

#### 4 Заключение

Хотя гиперконвергенцию можно рассматривать как естественную эволюцию идеи конвергенции, пока что оба подхода к организации инфраструктуры ЦОД, CI и HCI развиваются параллельно. Оба не идеальны, их очевидные слабые стороны:

- ограничение в выборе оборудования из-за единственного поставщика и, как следствие, возможно, более высокая стоимость по сравнению с традиционными решениями;
- трудности с балансировкой вычислительных и хранилищных ресурсов из-за их интеграции в типовых модулях;
- дополнительная нагрузка на внутреннюю сеть ЦОД вследствие переноса в нее хранилищного трафика.

Тем не менее оба подхода успешно завоевывают рынок благодаря их сильным сторонам:

- быстрому развертыванию и удобному масштабированию инфраструктуры;
- надежному обеспечению долговечности хранения данных за счет их распределения по инфраструктуре, особенно при реализации SDS;
- удобному управлению всеми компонентами с единой консоли;
- нетребовательности к обслуживающему персоналу.

Но наряду с этими часто повторяемыми общими для обоих инфраструктурных подходов достоинствами нельзя не отметить еще одно важное преимущество HCI, отличающее ее от CI. Для рядовых владельцев ЦОД покупка гиперконвергентной инфраструктуры может оказаться самым простым способом обретения полнофункционального программно-определяемого хранилища данных и, вероятно, единственно возможным — готовой программно-определяемой сети и даже, в перспективе, целиком программно-определяемого ЦОД.

#### Литература

1. Егоров В. Б. Некоторые вопросы практической реализации концепции SDN // Системы и средства информатики, 2016. Т. 26. № 1. С. 109–120. doi: 10.14357/08696527160108.

2. *Егоров В. Б.* Практичное определение «программно определяемого» // Системы и средства информатики, 2019. Т. 29. № 2. С. 85–94. doi: 10.14357/08696527190208.
3. *Ганьжа Д.* Гиперконвергенция: ИТ-инфраструктура на раз, два, три // Ж. сетевых решений / LAN, 2016. № 5. С. 33–39.
4. *Green J., Lowe S. D., Davis D. M.* The fundamentals of hyperconverged infrastructure. — ActualTech Media, 2018. 48 p. <https://www.actualtechmedia.com/wp-content/uploads/2018/04/SCALE-The-Fundamentals-of-Hyperconverged-Infrastructure-v2018.pdf>.
5. *Azeem S. A., Sharma S. K.* Study of converged infrastructure & hyper converged infrastructure as future of data centre // Int. J. Advanced Research Computer Science, 2017. Vol. 8. No. 5. doi: 10.26483/IJARCS.V8I5.3476.
6. *McMonigal C., Metz J., McDonald A.* Networking requirements for hyperconverged infrastructure (HCI). — SNIA, 2019. 37 p. <https://www.snia.org/sites/default/files/ESF/Networking-Requirements-for-Hyperconvergence-Final.pdf>.
7. *Подмоков Д.* Гиперконвергенция: стоит брать, но не стоит торопиться // Jetinfo, 2018. № 1-2. <https://www.jetinfo.ru/giperkonvergenziya-stoit-brat-no-ne-stoit-toropitsya>.
8. *Лукичев А.* Виртуализация нового типа // Jetinfo, 2014. № 11. <https://www.jetinfo.ru/virtualizatsiya-novogo-tipa>.
9. *Whiteley R.* Your next IT budget: 6 ways to support business growth // CIO Magazine, July 20, 2010. <https://www.cio.com/article/280069/your-next-it-budget-6-ways-to-support-business-growth.html>.
10. *Егоров В. Б.* Некоторые вопросы программного определения хранилища данных // Системы и средства информатики, 2021. Т. 31. № 2. С. 70–79.
11. *Crump G.* Hyperconverged 101 — the importance of SDN in HCI // StorageSwiss, 2019. <https://storageswiss.com/2019/05/08/hyperconverged-101-the-importance-of-sdn-in-hci/>.
12. *Sheldon R.* How software-defined networking layers work with HCI // TechTarget, April 9, 2018. <https://searchconvergedinfrastructure.techtarget.com/tip/How-software-defined-networking-layers-work-with-HCI>.
13. *Егоров В. Б.* К вопросу о соотношении программно определяемых и традиционных IP-сетей // Системы и средства информатики, 2022. Т. 32. № 1. С. 73–82. doi: 10.14357/08696527220107.

Поступила в редакцию 22.11.22

---

---

## THE SOFTWARE-DEFINED NETWORKING IN CONVERGED AND HYPERCONVERGED INFRASTRUCTURES

*V. B. Egorov*

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

**Abstract:** The converged (CI) and hyperconverged (HCI) infrastructures are widely discussed now; nevertheless, the terms CI and HCI themselves have not



obtained a generally accepted definition. In conditions of uncertainty with the interpretation of the terms, this article attempts to clarify the role of the network in such infrastructures as well as to evaluate the perspective of implementing their internal networks on the software-defined networking (SDN) principles. The centralized control of all infrastructural components from a single console as a distinctive feature of the CI meets the basic SDN principles. However, this condition is necessary but insufficient and a typical CI network can only be considered as a step towards the SDN. A full-fledged implementation of SDN becomes feasible in the HCI, where the HCI supplier is actually forced, due to hyperconvergence peculiarities, to implement in his product a software-defined storage (SDS) and is able to supplement the product with an SDN network. For ordinary data center owners, the purchase of a hyperconverged infrastructure may be not only the easiest way to acquire an effective SDS but almost the only feasible one to get a ready SDN.

**Keywords:** converged infrastructure; hyperconverged infrastructure; software-defined networking; CI; HCI; SDN

**DOI:** 10.14357/08696527230110

## References

1. Egorov, V. B. 2016. Nekotorye voprosy prakticheskoy realizatsii kontseptsii SDN [Some issues of the SDN concept practical implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(1):109–120. doi: 10.14357/08696527160108.
2. Egorov, V. B. 2019. Praktichnoe opredelenie “programmno opredelyaemogo” [A practical definition of “software-defined”]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):85–94. doi: 10.14357/08696527190208.
3. Ganzha, D. 2016. Giperkonvergentsiya: IT-infrastruktura na raz, dva, tri [Hyperconvergence: IT infrastructure on once, two, three]. *Zh. setevykh resheniy/LAN [J. Network Solutions LAN]* 5:33–39.
4. Green, J., S. D. Lowe, and D. M. Davis. 2018. *The fundamentals of hyperconverged infrastructure*. ActualTech Media. 48 p. Available at: <https://www.actualtechmedia.com/wp-content/uploads/2018/04/SCALE-The-Fundamentals-of-Hyperconverged-Infrastructure-v2018.pdf> (accessed April 12, 2023).
5. Azeem, S. A., and S. K. Sharma. 2017. Study of converged infrastructure & hyper converge infrastructure as future of data centre. *Int. J. Advanced Research Computer Science* 8(5):900–903. doi: 10.26483/IJARCS.V8I5.3476.
6. McMonigal, C., J. Metz, and A. McDonald. 2019. *Networking requirements for hyperconverged infrastructure*. SNIA. 37 p. Available at: <https://www.snia.org/sites/default/files/ESF/Networking-Requirements-for-Hyperconvergence-Final.pdf> (accessed November 16, 2022).
7. Podmokov, D. 2018. Giperkonvergentsiya: stoit brat', no ne stoit toropit'sya [Hyperconvergence: It's worth taking, but don't rush it]. *Jetinfo* 1-2. Available at: <https://www.jetinfo.ru/giperkonvergentsiya-stoit-brat-no-ne-stoit-toropitsya/> (accessed November 16, 2022).
8. Lukichev, A. 2014. Virtualizatsiya novogo tipa [Virtualization of a new type]. *Jetinfo* 11. Available at: <https://www.jetinfo.ru/virtualizatsiya-novogo-tipa/> (accessed November 16, 2022).



9. Whiteley, R. July 20, 2010. Your next IT budget: 6 ways to support business growth. *CIO Magazine*. Available at: <https://www.cio.com/article/280069/your-next-it-budget-6-ways-to-support-business-growth.html> (accessed April 12, 2023).
10. Egorov, V. B. 2021. Nekotorye voprosy programmno opredeleniya khranilishcha dannykh [Some issues of software-defined data centers]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):70–79. doi: 10.14357/08696527200210.
11. Crump, G. 2019. Hyperconverged 101 — the importance of SDN in HCI. *StorageSwiss*. Available at: <https://storageswiss.com/2019/05/08/hyperconverged-101-the-importance-of-sdn-in-hci/> (accessed November 16, 2022).
12. Sheldon, R. April 9, 2018. How software-defined networking layers work with HCI. *TechTarget*. Available at: <https://searchconvergedinfrastructure.techtarget.com/tip/How-software-defined-networking-layers-work-with-HCI> (accessed November 16, 2022).
13. Egorov, V. B. 2022. K voprosu o sootnoshenii programmno opredelyaemykh i traditsionnykh IP-setey [Interrelation between the software-defined and conventional IP-networks]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):73–82. doi: 10.14357/08696527220107.

*Received November 22, 2022*

## **Contributor**

**Egorov Vladimir B.** (b. 1948) — Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; VEgorov@ipiran.ru

## ЦЕНТРЫ КОМПЕТЕНЦИИ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ И НАЦИОНАЛЬНАЯ ТЕХНОЛОГИЧЕСКАЯ ИНИЦИАТИВА

*А. П. Сучков*<sup>1</sup>

**Аннотация:** Задача достижения технологического суверенитета может быть решена путем консолидации усилий на государственном уровне с выделением приоритетных направлений и задач развития инновационных технологий и определением путей их достижения. В связи с этим автономной некоммерческой организации «Агентство стратегических инициатив по продвижению новых проектов» рекомендовано подготовить стратегический план развития Национальной технологической инициативы (НТИ) на долгосрочный период и предложения по контролю за его соблюдением. В статье рассматриваются вопросы формирования целостной картины совокупности функций и задач НТИ в интересах данного стратегического плана и применительно к проблематике технологий искусственного интеллекта (ИИ).

**Ключевые слова:** Национальная технологическая инициатива; центры компетенций; технологии искусственного интеллекта

**DOI:** 10.14357/08696527230111

### 1 Введение

Создание и государственная поддержка Центров компетенций (ЦК) в рамках НТИ реализуется в соответствии с постановлением Правительства Российской Федерации [1]. Цель программы их поддержки — сформировать сеть инженерно-образовательных консорциумов на базе российских университетов и научных организаций для создания инновационных решений в области «сквозных» технологий, обеспечивающих глобальное лидерство компаниям, которые используют данные технологии для производства продуктов и услуг. На конкурсной основе в 2017–2021 гг. был создан 21 Центр компетенций НТИ, в 2022 г. планируется провести конкурсный отбор еще 3 Центров.

Центры создаются в партнерстве с российскими и зарубежными организациями (университетами и ведущими научными организациями, коммерческими партнерами), которые участвуют в постановке исследовательских задач. Работа ЦК НТИ осуществляется в соответствии с сетевыми принципами: совместная работа распределенного коллектива, включая партнеров по коммерциализации технологий.

---

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ASuchkov@ipiran.ru

Основными показателями эффективности работы Центров компетенций НТИ являются численность подготовленных специалистов, объем доходов от приносящей доход деятельности, источником которых служит деятельность центра, и количество лицензированных технологий, т. е. непосредственный вклад в развитие экономики страны.

Отбор получателей грантов осуществляется на конкурсной основе. Функциями оператора конкурсного отбора в соответствии с постановлением Правительства РФ № 1119 от 05.07.2021 наделен Фонд поддержки проектов НТИ, который осуществляет процедуры конкурсного отбора, контрактации, сопровождения и мониторинга деятельности Центров компетенций НТИ. До конца 2024 г. на поддержку Центров компетенций НТИ предусмотрено государственное финансирование в размере 16 млрд руб. При этом модель финансирования Центров предусматривает поэтапное замещение бюджетных грантовых денег средствами софинансирования.

Определены и задачи Центров компетенций НТИ [2]:

- **прямая трансляция результатов фундаментальной науки в инженерные приложения.** Междисциплинарные исследовательские программы обеспечивают «переложение» фундаментальных научных результатов и идей через прикладные исследования и разработки в конкретные технологии в интересах конкретных промышленных партнеров;
- **технологический трансфер через кооперацию с промышленными партнерами.** Центры компетенций НТИ формируют устойчивую связь между академической сферой (университеты, научные организации) и промышленными партнерами;
- **подготовка лидеров разработки новых технологий через реализацию образовательных программ.** Центры компетенций НТИ создают и реализуют образовательные программы инженерного профиля. Для студентов этих программ предусматривается обязательное участие в исследовательской деятельности ЦК НТИ в форме непосредственной работы над проектами в командах с представителями промышленных партнеров.

Как видно, построение системы ЦК НТИ осуществляется снизу вверх — от существующих заделов научных организаций (иначе конкурс выиграть невозможно). Однако задача достижения технологического суверенитета может быть решена только путем консолидации усилий на государственном уровне с выделением приоритетных направлений и задач развития инновационных технологий и определением путей их достижения. В связи с этим автономной некоммерческой организации «Агентство стратегических инициатив по продвижению новых проектов» рекомендовано подготовить стратегический план развития НТИ на долгосрочный период и предложения по контролю за его соблюдением [3].

Следует отметить, что ряд централизованных функций НТИ, таких как разработка и реализация образовательных программ высшей школы, направленных на формирование компетенций в этой области, ведение реестра проектов [1, 3],

заложены в нормативно-правовые документы, однако имеется необходимость формирования целостной картины совокупности этих функций и задач НТИ в интересах данного стратегического плана и применительно к проблематике технологий ИИ.

## **2 Стратегические цели и задачи Национальной технологической инициативы в области искусственного интеллекта**

Формирование стратегического плана развития НТИ в области ИИ необходимо осуществлять с учетом Национальной стратегии развития ИИ на период до 2030 г. [4] в части технологического суверенитета Российской Федерации в области ИИ, включая обеспечение необходимого уровня самодостаточности и преимущественного использования отечественных технологий и технологических решений. При этом необходимо учитывать угрозы в этой области, такие как неразвитость нормативной и правовой базы, отставание в части системного и прикладного программного обеспечения, отсутствие высокопроизводительных отечественных средств вычислительной техники (СВТ), недостаток специалистов.

В Комплексной научно-технической программе «Искусственный интеллект как драйвер цифровой трансформации экономики России» (2019), разработанной ФИЦ ИУ РАН [5], в качестве базовых выделены следующие технологии:

- технологии машинного обучения;
- технологии извлечения знаний из различных источников;
- технологии интеллектуального анализа больших данных;
- технологии распознавания образов;
- технологии прогнозирования и поддержки принятия решений;
- технологии планирования и управления целенаправленным поведением в неструктурированных средах;
- когнитивные технологии;
- технологии мультиагентного управления и диспетчирования ресурсов в распределенных системах;
- технологии мультимодальной аналитики и рассуждений;
- технологии обработки естественных языков;
- технологии автономных непилолируемых интеллектуальных систем;
- технологии создания машин знаний и их операционных систем;
- технологии гибридного человеко-машинного интеллекта.

Центры компетенций НТИ, связанные с проблематикой ИИ

Сквозная технология	Победитель конкурсного отбора	Наименование Центра
Искусственный интеллект	МФТИ	Центр НТИ по направлению «Искусственный интеллект»
Нейротехнологии, технологии виртуальной и дополненной реальности	ДВФУ	Центр НТИ по направлению «Нейротехнологии, технологии виртуальной и дополненной реальности»
Технологии хранения и анализа больших данных	МГУ имени М. В. Ломоносова	Центр технологий хранения и анализа больших данных
Технологии компонентов робототехники и мехатроники	Университет Иннополис	Центр технологий компонентов робототехники и мехатроники
Технологии сенсорики	МИЭТ	Центр НТИ МИЭТ «Сенсорика»
Технологии машинного обучения и когнитивные технологии	ИТМО	Национальный центр когнитивных разработок

Часть из этих технологий ИИ стали направлениями деятельности ряда ЦК НТИ (см. таблицу).

В качестве стратегических целей, реализуемых Фондом поддержки проектов НТИ, предлагается:

- создание условий для ускоренного и скоординированного развития отечественного программно-аппаратного, информационного и кадрового потенциала в области всех базовых технологий ИИ на основе формирования консорциума ЦК и промышленных партнеров с целью преодоления технологического отставания в этой области;
- сокращение сроков и снижение стоимости реализации дорожных карт создания систем ИИ с элементами ИИ за счет унификации, стандартизации и исключения дублирования;
- создание механизмов учета и формирования компетенций: актуальных задач и лучших решений в данной области, учет потребностей отечественных разработчиков систем ИИ на основе создания соответствующих реестров.

Достижение приведенной выше совокупности целей будет осуществляться на основе решения комплекса задач в нормативно-правовой, информационно-технологической, организационной и кадровой сферах.

1. В нормативной правовой сфере:

- обеспечение нормативной и правовой базы для создания и функционирования организационных структур ЦК НТИ;

- нормативно-техническое обеспечение в части создания и совершенствования стандартов и критериев.
2. В части аппаратного обеспечения:
- мониторинг отечественных СВТ для целей НТИ;
  - формирование горизонта перспективных СВТ на основе анализа перспективных задач и учета потребностей разработчиков систем ИИ.
3. В части развития программного обеспечения:
- формирование перечня верифицированного системного программного обеспечения;
  - формирование доверенных библиотек и пакетов инструментальных программных средств разработки систем ИИ, в том числе с открытым кодом;
  - формирование горизонта перспективного программного обеспечения на основе анализа перспективных задач и учета потребностей разработчиков систем ИИ;
  - продвижение «прозрачных» (гибридных) технологий ИИ, основанных как на обучаемых алгоритмах (в частности, нейросетях глубокого обучения), так и логическом выводе.
4. В части развития информационного обеспечения:
- разработка и учет верифицированных массивов данных для обучения и тестирования систем ИИ;
  - формирование горизонта перспективного информационного обеспечения на основе анализа перспективных задач и учета потребностей разработчиков систем ИИ.
5. Основные задачи в части организационного обеспечения:
- формирование организационной структуры для управления процессами создания и обеспечения функционирования ЦК НТИ;
  - проведение аудита программно-аппаратных средств и аналитических систем ИИ заинтересованных органов и организаций с целью обобщения потребностей и опыта их разработки;
  - систематизация методов ИИ;
  - создание Фонда аналитических технологий ИИ, предназначенного для учета существующих и перспективных технологий ИИ, а также формирования базы данных (БД) актуальных научно-технических проблем и практических задач в данной области, учета потребностей отечественных разработчиков систем ИИ;

- на основе анализа потребностей промышленных партнеров и академической сферы формирование обоснованных требований по разработке программно-аппаратного, информационного обеспечения и формированию кадрового потенциала в указанной области.

6. Основные задачи в части кадрового обеспечения:

- формирование необходимых научно-исследовательских компетенций и технологических практик в сфере создания и функционирования систем ИИ, включая оценку кадровых потребностей;
- разработка специализированных программ повышения квалификации и переподготовки в этой области;
- подготовка (повышение квалификации, переподготовка) лиц, задействованных в создании и обеспечении функционирования ЦК НТИ.

### **3 Перспективные функциональные возможности Фонда поддержки проектов Национальной технологической инициативы в области искусственного интеллекта**

Стратегический план развития НТИ должен обеспечить последовательную реализацию стратегических целей и задач по нейтрализации существующих угроз и рисков в области разработки и функционирования систем ИИ на основе учета, аккумуляции существующих компетенций и ресурсов в области разработки систем ИИ, формирования дорожных карт развития, а также контроль их выполнения.

Рассмотрим основные функциональные возможности Фонда поддержки проектов НТИ в области ИИ.

1. *Учет и мониторинг существующих и перспективных угроз реализации систем ИИ.*

Негативные воздействия на ход формирования и реализации задач НТИ могут оказывать как внешние, так и внутренние факторы на всех стадиях жизненного цикла систем ИИ [6]. При этом осуществляется учет рисков и угроз, связанных как с разработкой систем ИИ, так и с их применением. К внутренним факторам рисков относятся те, которые обуславливаются процессами организации функционирования НТИ. Внешние факторы риска характеризуют внешнее отрицательное воздействие на стадиях формирования требований, реализации задач НТИ и ее функционирования. Внутренние причины — конфликт интересов сторон, некомпетентность заказчика и исполнителя, отсутствие независимой экспертной оценки — могут привести к некорректному целеполаганию. Учет внешних источников угроз позволяет ослабить риски невозможности выполнения требований и реализации решений в условиях технологического отставания и внешних ограничений

(санкции) и организационных проблем, связанных с нормативно-правовой базой и отсутствием независимого контроля.

Наиболее адекватной моделью представления данных мониторинга угроз считается динамическая семантическая сеть (ДСС) [7]. Динамическая семантическая сеть определяется как множество узлов (в данном случае это информационные объекты предметной области) и множество связей (отношений). Информационные объекты семантической сети обладают атрибутами, один из основных среди которых — «состояние объекта», а также идентификационными характеристиками, позволяющими их однозначно определять.

Вопросы применения технологий ИИ порождают множество новых рисков и угроз, особенно в военной сфере. Они связаны с ошибками в архитектуре, злонамеренных действиях разработчиков и других субъектов процесса, недостаточностью данных для обучения и тестирования и другими причинами. Это приводит к спектру труднопрогнозируемых угроз и рисков, последствия которых сегодня не могут быть достоверно оценены. Среди них отмечаются программная враждебность, пассивный риск, случайный риск, риск непостижимости (см., например, [8]).

Другие риски систем ИИ могут быть связаны с их разработкой и применением в военной области. Учет и мониторинг данных рисков позволяет уточнять перечень перспективных разработок систем ИИ. Как пример можно рассмотреть поставки Великобританией Украине необитаемых подводных аппаратов REMUS 100 (Remote Environmental Measuring Units 100 m), которые представляют собой совместную разработку норвежской компании Hydroid, входящей в состав Kongsberg Maritime (подводная платформа), и британского института Woods Hole Oceanographic Institution (комплекс бортовой аппаратуры, основанный на технологиях ИИ) [9].

На рис. 1 представлен фрагмент ДСС потенциальной угрозы национальной безопасности, связанной с технологиями ИИ. Дальнейшими целями мониторинга этой угрозы могут стать оценка способов реализации этой угрозы, оценка связанного с ними ущерба и меры возможного противодействия.

## 2. Информационный мониторинг мирового уровня разработок ИИ на основе анализа открытых источников.

Для целей мониторинга необходимо сформировать перечень источников данных, представляющих собой зарубежные БД научно-технической информации:

- публикации на английском языке ресурса arXiv.org (2 млн документов) по естественным наукам;
- англоязычная Википедия (более 5 млн статей);
- американские патенты (USPTO с 2002 по 2022 гг.): более 3 млн патентов;
- международные патенты (WIPO): 3 млн патентов.



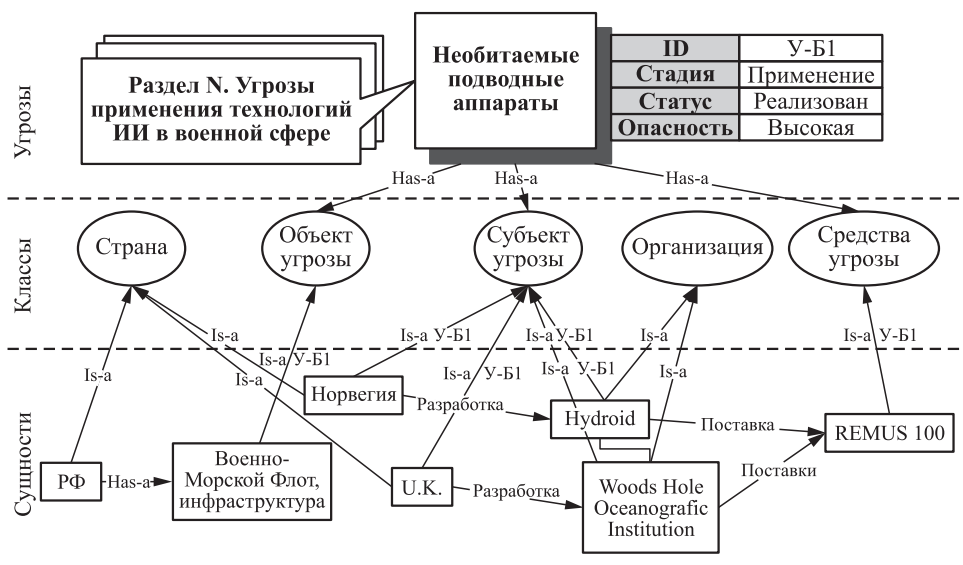


Рис. 1 Фрагмент ДСС угрозы военного применения систем ИИ

Важнейшим источником данных служат зарубежные электронные реферативные БД.

### 3. Создание Фонда аналитических технологий ИИ.

Фонд предназначен для учета существующих и перспективных технологий ИИ, а также формирования БД актуальных научно-технических проблем и практических задач в данной области, учета потребностей отечественных разработчиков систем ИИ. С целью его создания осуществляется:

- аудит, инвентаризация и учет существующих отечественных программно-аппаратных, информационных и кадровых ресурсов, необходимых для целей НТИ;
- мониторинг и учет актуальных проблем и задач ИИ для обеспечения потребностей промышленных партнеров;
- формирование перечня необходимых научно-исследовательских компетенций и технологических практик в сфере создания и функционирования систем ИИ, включая оценку кадровых потребностей;
- мониторинг и учет потребностей разработчиков в ресурсах (программно-аппаратных и др.) для разработки систем ИИ.

Аудит можно осуществлять как по материалам организаций, участвующих в конкурсах, так и путем целенаправленного сбора данных с использованием, например, имеющегося опыта ФИЦ ИУ РАН [10, 11] (рис. 2).

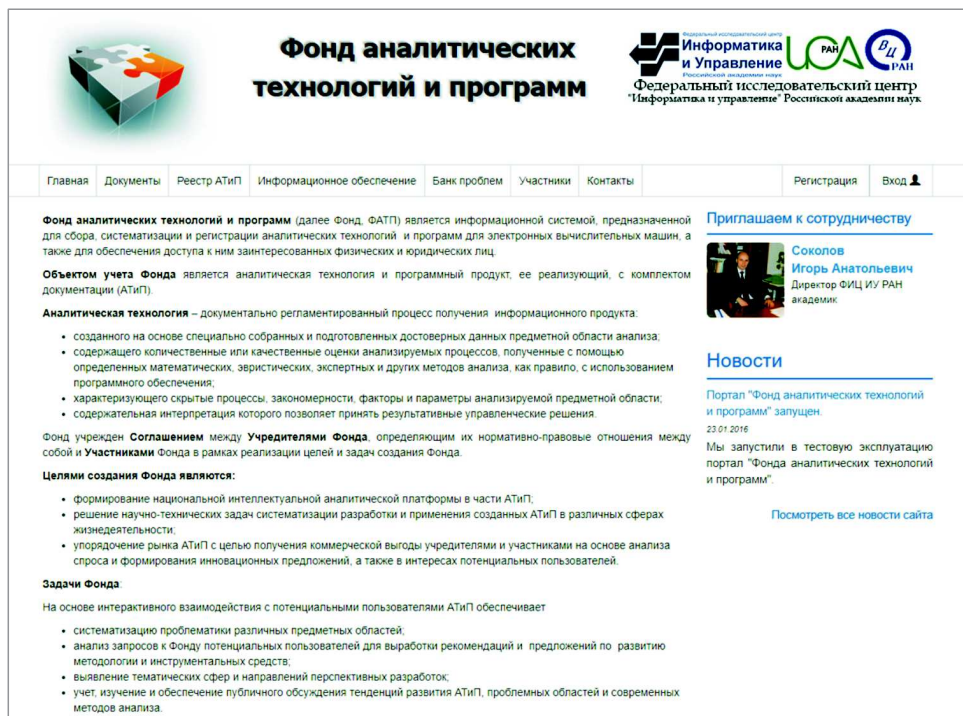


Рис. 2 Веб-сайт Фонда аналитических технологий и программ

Это должно лечь в основу:

- формирования национальной интеллектуальной аналитической платформы;
- решения научно-технических задач систематизации разработки и применения созданных систем ИИ в различных сферах жизнедеятельности;
- упорядочения рынка технологий ИИ с целью получения коммерческой выгоды участниками процесса на основе анализа спроса и формирования инновационных предложений, а также в интересах промышленных партнеров.

#### 4. Формирование и корректировка актуального комплекса требований к ресурсам технологий ИИ:

- требований и методических материалов в части аппаратного обеспечения технологий ИИ;
- требований к отечественной операционной системе для суперкомпьютерных вычислений;

- требований к отечественным инструментальным средствам для разработки программных компонентов средств ИИ;
- разработка верифицированных наборов обучающих данных;
- требований к системе образования для обеспечения перспективных кадровых потребностей.

5. *Функции, связанные с планированием деятельности и контролем исполнения:*

- формирование комплекса дорожных карт с целью реализации решений по созданию и развитию Центров компетенции НТИ;
- мониторинг процессов деятельности ЦК НТИ по реализации дорожных карт;
- тестирование всех видов ресурсов и систем на способность выполнять возложенные на них задачи с требуемым качеством.

## 4 Заключение

При формировании стратегического плана развития НТИ необходимо всемерно поддерживать процессы достижения технологического суверенитета путем консолидации усилий на государственном уровне с выделением приоритетных направлений и задач развития инновационных технологий и определением путей их достижения. В связи с этим актуальными становятся задачи создания механизмов учета и формирования компетенций: актуальных задач и лучших решений в данной области, учет потребностей отечественных разработчиков систем ИИ на основе создания соответствующих реестров, а также мониторинга состояния данной предметной области во всем ее многообразии.

## Литература

1. Постановление Правительства Российской Федерации от 16 октября 2017 г. № 1251. <http://static.government.ru/media/files/V2pqqEl6LX8f1KM3GduslGVAvzuyAUoy.pdf>.
2. Фонд НТИ: поддержка Центров компетенции НТИ. <https://nti.fund/support/centers>.
3. О реализации Национальной технологической инициативы: Постановление Правительства РФ от 18.04.2016 № 317 (ред. от 13.07.2022).
4. Национальная стратегия развития искусственного интеллекта на период до 2030 года: Указ Президента РФ от 10 октября 2019 г. № 490.
5. Комплексная научно-техническая программа (КНТП) «Искусственный интеллект как драйвер цифровой трансформации экономики России» ФИЦ ИУ РАН. — Совет по приоритетным направлениям научно-технологического развития, 2019 (неопубл.).

6. Сучков А. П. Классификация уязвимостей интегрированных систем управления на ранних стадиях жизненного цикла // Системы и средства информатики, 2017. Т. 27. № 4. С. 132–143. doi: 10.14357/08696527170410.
7. Сучков А. П. Информационная структура угроз национальной безопасности // Системы и средства информатики, 2017. Т. 27. № 2. С. 113–124. doi: 10.14357/08696527170210.
8. Насколько опасен искусственный интеллект. <https://pctune.ru/iskusstvennyi-intellekt.html>.
9. Британские безэкипажные подводные аппараты для Украины. <https://topwar.ru/202052-britanskie-bezejekipazhnye-podvodnye-apparaty-dlja-ukrainy.html>.
10. Зацаринный А. А., Соченков И. В., Сучков А. П. Некоторые научно-методические вопросы создания Фонда аналитических технологий и программ // Искусственный интеллект и принятие решений, 2016. № 2. С. 78–82.
11. Фонд аналитических технологий и программ. <https://fatp.frccsc.ru>.

*Поступила в редакцию 22.11.22*

---

---

## COMPETENCE CENTERS FOR ARTIFICIAL INTELLIGENCE AND THE NATIONAL TECHNOLOGY INITIATIVE

*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 task of achieving technological sovereignty can be solved by consolidating efforts at the state level, highlighting priority areas and tasks for the development of innovative technologies and identifying ways to achieve them. In this regard, the autonomous nonprofit organization “Agency for Strategic Initiatives for the Promotion of New Projects” is recommended to prepare a strategic plan for the development of the National Technology Initiative (NTI) for the long term and proposals for monitoring its compliance. The article discusses the issues of forming a holistic picture of the totality of the functions and tasks of the NTI in the interests of this strategic plan and in relation to the problems of artificial intelligence technologies.

**Keywords:** National Technology Initiative; competence centers; artificial intelligence technologies

**DOI:** 10.14357/08696527230111

### References

1. Postanovlenie Pravitel'stva Rossiyskoy Federatsii ot 16 oktyabrya 2017 g. No.1251 [Decree of the Government of the Russian Federation No.1251 dated 16.10.2017]. Available at: <http://static.government.ru/media/files/V2pqqE16LX8f1KM3GduslGVAvzuyAUoy.pdf> (accessed March 3, 2023).

2. Fond NTI: Podderzhka Tsentrov kompetentsii NTI [NTI Foundation: Support of NTI Competence Centers]. Available at: <https://nti.fund/support/centers/> (accessed March 3, 2023).
3. O realizatsii Natsional'noy tekhnologicheskoy initsiativy: Postanovlenie Pravitel'stva RF ot 18.04.2016 No. 317 (red. ot 13.07.2022) [On the implementation of the National Technological Initiative: Decree of the Government of the Russian Federation No. 317 dated 18.04.2016 (edition of 13.07.2022)].
4. Natsional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda: Ukaz Prezidenta RF ot 10 oktyabrya 2019 g. No. 490 [National Strategy for the Development of Artificial Intelligence for the period up to 2030: Decree of the President of the Russian Federation No. 490 dated 10.10.2019].
5. FITZ IU RAN. 2019 (unpubl.). Kompleksnaya nauchno-tekhnicheskaya programma (KNTP) "Iskusstvennyy intellekt kak drayver tsifrovoy transformatsii ekonomiki Rossii" [Integrated Scientific and Technical Program (ISTP) "Artificial Intelligence as a Driver of Digital Transformation of the Russian Economy"].
6. Suchkov, A. P. 2017. Klassifikatsiya uyazvimostey integrirovannykh sistem upravleniya na rannikh stadiyakh zhiznennogo tsikla [Classification of vulnerabilities of integrated control systems at the early stages of life cycle]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):132–143. doi: 10.14357/08696527170410.
7. Suchkov, A. P. 2017. Informatsionnaya struktura ugroz natsional'noy bezopasnosti [The information structure of threats to national security] *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):113–124. doi: 10.14357/08696527170210.
8. Naskol'ko opasen iskusstvennyy intellekt [How dangerous is artificial intelligence]. Available at: <https://pctune.ru/iskusstvennyi-intellekt.html> (accessed March 3, 2023).
9. Britanskiye bezekipazhnye podvodnye apparaty dlya Ukrainy [British unmanned underwater vehicles for Ukraine]. Available at: <https://topwar.ru/202052-britanskiye-bezjekipazhnye-podvodnye-apparaty-dlja-ukrainy.html> (accessed March 3, 2023).
10. Zatsarinny, A. A., I. V. Sochenkov, and A. P. Suchkov. 2016. Nekotoryye nauchno-metodicheskiye voprosy sozdaniya Fonda analiticheskikh tekhnologiy i programm [Some scientific and methodological issues during the creation of the analytical technologies and programs foundation]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making] 2:78–82.
11. Fond analiticheskikh tekhnologiy i programm [Foundation for Analytical Technologies and Programs]. Available at: <https://fatp.frccsc.ru/> (accessed March 3, 2023).

*Received November 22, 2022*

## Contributor

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

## ТЕОРИЯ S-СИМВОЛОВ: КОНЦЕПТУАЛЬНЫЕ ОСНОВАНИЯ

*В. Д. Ильин*<sup>1</sup>

**Аннотация:** Предложенная теория S-символов представляет собой расширенное обобщение теории S-моделирования и 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 (ИТ) различного назначения. В статье представлена первая часть (из четырех) описания теории. Приведено обоснование целесообразности и основные понятия (S-символ, S-код, S-сигнал и др.). Определены виды и типы S-(символов, кодов и сигналов). Введены отношения эквивалентности, порядка и принадлежности, задаваемые на системах S-(символов, кодов и сигналов). Определения сопровождаются примерами.

**Ключевые слова:** теория S-символов; S-символ; S-код; S-сигнал; S-среда; искусственный интеллект; информационная технология

**DOI:** 10.14357/08696527230112

### 1 Введение

В 1975 г. Аллен Ньюэлл (Allen Newell) и Герберт Саймон (Herbert A. Simon) опубликовали лекцию, посвященную концептуальному обеспечению построения интеллектуальных систем [1]. Первый раздел этой работы посвящен *символам и системам физических символов* (Symbols and Physical Symbol Systems). *Физический* означает *подчиняющийся законам физики и пригодный для реализации в виде инженерных систем*. В своей гипотезе Ньюэлл и Саймон утверждают, что *система физических символов имеет необходимые и достаточные средства для общего разумного действия*. В рассуждениях об интеллектуальных системах Ньюэлл и Саймон отмечают первостепенную важность эффективного символьного представления сущностей.

*Теория S-символов*<sup>2</sup> служит расширенным обобщением *теории S-моделирования* [2]. Понятие *S-моделирования* было определено в 1989 г. на страницах 248–249 книги [3], посвященной методологии автоматизированного конструирования программных систем. В 2007–2017 гг. в Институте

---

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

<sup>2</sup>Далее префикс S означает принадлежность понятий теории S-символов.

проблем информатики РАН под научным руководством автора был выполнен ряд научно-исследовательских работ, посвященных методологии S-моделирования и ее применению. В 2008–2015 гг. обучающимся на базовой кафедре Института проблем информатики РАН в Российском технологическом университете (МИРЭА) читался курс лекций *Символьное моделирование в информатике*. Методологические основания этого курса представлены в статье [4].

**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) [5].

*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) □ ≈ определение (здесь и далее символ ≈ заменяет слово «означает»);

◇ (фрагмент описания) ◇ ≈ замечание;

○ (фрагмент описания) ○ ≈ пример.

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

## 2 Концептуальные основания

Изучая мир и себя, люди строят *символьные* (аудио-, видео- и др.) *модели сущностей*, отражающие изучаемые объекты и связи между ними. Эти модели [○ представленные языками взаимодействия, спецификации и программирования; *цифровыми двойниками* различных объектов [6–9] и др. ○] служат не только средством познания, но и инструментарием, помогающим изобретать искусственные вещи, расширяющие и совершенствующие естественные возможности человека. Такой инструментарий принято относить к арсеналу *искусственного интеллекта*.

В наши дни люди и созданная ими техника действуют в *символьно-кодОВО-сигнальной* среде (*S-среде*), которая в современной реализации является *цифровой*. Совершенствованию этой среды посвящены труды исследователей и инженеров.

□ ***S-(символы, коды, сигналы)*** и ***S-символьные конструкции*** (○ языки спецификации, программы управления поведением *S-машин* и др. ○) — *S-объекты, свойства, взаимосвязи и методы построения которых изучаются в теории S-символов.* □



S-объект рассматривается как отображение некоторого предмета изучения в S-среду, выполненное при заданных ограничениях. Не предполагается никаких ограничений на природу и масштабы отображаемых предметов. Они могут быть природными или созданными людьми, могут иметь любую физическую сущность и масштабы.

□ **S-символ** — заменитель природного или изобретенного объекта, обозначающий этот объект и служащий элементом определенной системы построения *S-сообщений* (○ видео-, аудиосообщений или др. ○) в S-среде [5], рассчитанных на восприятие человеком или S-машиной. □

○ Текст электронного документа, электронная географическая карта, видеоролик — все это *S-сообщения*, файлы S-кодов которых хранятся на накопителях (○ SSD, жестком диске или др. ○) S-машин. ○

○ *Шрифт Брайля* для слепых — *система фактурных S-символов* для построения текстовых сообщений, рассчитанных на восприятие осязанием пальцами рук; нотное письмо, *система нотных s-символов* — средство построения музыкальных аудиосообщений, представленных в графической форме, а *система S-символов шахматной нотации* — средство записи шахматных партий в виде текстовых сообщений. Русский алфавит вместе со знаками препинания — *система текстовых S-символов для построения сообщений по правилам грамматики русского языка* (а каждый элемент алфавита является заменителем звука, применяемого в речевых сообщениях). ○

□ **S-сообщение** — конечная упорядоченная (в соответствии с некоторыми правилами) совокупность S-символов или ее S-код в S-сигнальном представлении. Рассчитано на передачу, прием, распознавание и интерпретацию получателем (человеком или S-машиной). □

В компьютерах, смартфонах и других S-машинах S-символы представлены в форме S-кодов. Отличительным признаком S-символьных объектов, существующих в S-среде (○ файлов электронных книг, статей, видеозаписей, карт, компьютерных программ и др. ○), служат легкоосуществимые копирование без искажений, распространение и хранение копий (по сравнению с несимвольными объектами: физическими моделями, макетами научно-технических сооружений и др.).

Определены **виды S-символов**, для каждого из которых — *типы S-символов*. □ *Виду S-символов* соответствует *средство приема сообщений*, которым наделен принимающий сообщение человек (или робот): *визуальный* — зрение; *аудио* — слух; *тактильный* — осязание; *запаховый* — обоняние. □

◇ В современной (начала 2020-х гг.) S-среде сравнительно продуктивно используются зрение (восприятие текста, неподвижных и подвижных изображений и др.) и слух (восприятие речи, музыки и др.). Осязание используется в мобильных телефонах (для приема вибровызова), в игровых устройствах и др.; обоняние — в стадии экспериментирования (выпускаются приборы для распознавания запахов). ◇



Для производства S-символьных сообщений человек может использовать органы речи, части тела, производящие различные движения (жесты рук, движения пальцев и др.), и глаза (выбор объекта, указанием на место его нового размещения и др.).

◇ В современной S-среде относительно продуктивно используются движения пальцев рук (посредством компьютерной клавиатуры, мыши, стилуса, сенсорного экрана и др.), распознавание звуковых команд и извещений находится в стадии активного становления, а использование движений глаз (как средства производства сообщений, направляемых S-машинам) — в стадии исследований. ◇

□ **Типу S-символов** соответствует множество символов, для которых определены набор атрибутов и семейство допустимых операций. □

*Виду S-визуальный* соответствуют типы: *S-графический* [для построения сообщений, содержащих неподвижные изображения (фотографии, схемы и др.)]; *видео* (для построения сообщений, содержащих подвижные изображения) и др.

○ *Таблица Менделеева* (на веб-странице) — *S-графическое представление сообщения о периодическом законе*, установившем зависимость свойств химических элементов от их атомной массы. ○

*Виду S-аудио* соответствуют типы: *S-речевой* (для построения сообщений, содержащих речевые фрагменты); *S-музыкальный* и др.

*Виду S-тактильный* соответствуют типы: *S-кинетический* [служит для формирования сообщений путем перемещения элементов устройств, предназначенных для интерфейса с S-машиной (○ вибровывоз мобильного телефона — сообщение о поступившем запросе на связь ○)]; *S-фактурный* [для передачи сообщений путем изменения фактуры поверхности элементов устройств (○ плоская → волнистая или ребристая и т. д. ○)]; *S-термический* (для передачи сообщений путем изменения температуры поверхности элементов интерфейсных устройств).

**Элементарные и составные S-символы.** □ S-символ любого типа, используемый для построения составных S-символов, называем *элементарным S-символом*. □ ○ Пиксель — элементарный графический S-символ растровых графических устройств. Пиктограмма — составной графический S-символ. ○

**S-отношения.** □ *S-эквивалентность* определяет взаимозаменяемость S-символов. Отношение эквивалентности может быть задано для любого числа S-символов (не менее двух). □ ○ Вибровывоз мобильного телефона (сообщение, составленное из тактильных S-символов) эквивалентен любому из аудиовывозов (○ сообщений, составленных из музыкальных S-символов ○). ○

□ *S-порядок* определяет последовательность S-символов. □ ○ Алфавит — последовательность текстовых S-символов. ○

□ *S-принадлежность* определяет вхождение S-символа в некоторый набор S-символов. □ ○ S-символ □, используемый в теории S-символов для выде-

ления утверждений, принадлежит набору S-символов *TSM-комплекса средств описания S-моделей*. ○

□ **S-код** служит заменителем S-символа, S-символьного сообщения, S-сигнала или S-сигнального сообщения. Используется для их представления в S-машинах. Предназначен для построения, сохранения, передачи, интерпретации сообщений и манипулирования ими в S-среде. □

◇ Цифровое кодирование S-символов и S-символьных сообщений позволяет применять методы решения задач, которые можно представить в виде программ, рассчитанных на выполнение цифровыми S-машинами. При цифровом кодировании S-символам ставят во взаимно однозначное соответствие числа, которые можно эффективно представить в памяти S-машины. При этом основание системы счисления выбирают равным числу устойчивых состояний, в которых могут находиться элементарные составляющие, из которых построена S-машина (○ в современных цифровых S-машинах используется двоичное представление кодов программ и данных, так как транзисторы имеют два устойчивых состояния ○).

Цифровое кодирование S-сигналов и S-сигнальных сообщений [○ с помощью аналого-цифровых преобразователей ○] — в основе построения современных систем связи, обработки данных и др. ◇

○ Графические S-символы планет *Солнечной системы* астрономы применяют при изготовлении электронных документов. При этом используются S-коды стандарта Юникод (The Unicode Standard) [○ планете *Меркурий* соответствует S-символ ♃ (S-код — U3+263F) ○]. ○

□ **S-сигнал** — физически реализованное представление S-символа (или S-символьного сообщения), S-кода (или S-кодированного сообщения), рассчитанное на передачу, прием, распознавание и интерпретацию S-машиной или человеком. □

○ В S-среде человек (или робот) получает S-сообщения, реализованные в виде оптических, звуковых и других S-сигналов. Технические устройства S-машин рассчитаны на работу с оптическими (○ цифровые фото- и видеокамеры ○), электрическими (○ микропроцессоры S-машин ○) и другими S-сигналами. ○

**S-объекты и их типизация.** □ **S-тип X** — множество X S-объектов, элементы которого имеют фиксированный набор атрибутов и семейство допустимых операций. Может иметь *S-подтипы*, называемые *специализациями S-типа X*, и *надтипы*, называемые *обобщениями S-типа X*. □

□ *Специализация S-типа X* — порождение *S-подтипа X* [::rule] (здесь двоенное двоеточие «::» — символ специализации) с семейством связей, расширенным добавлением связи *rule*. Выделяет подмножество  $X[::rule]$  множества X. Специализацией называем и результат  $X[::rule]$  этого порождения ( $X > X[::rule]$ ). □

○ Тип *S-текстовый* — специализация типа *S-графический*. ○

Специализацию S-типа, заданную последовательностью добавленных связей  $X[:(rule\ 1)::rule\ 2]$ , называем специализацией типа  $X[:(rule\ 1)]$  по связи  $rule\ 2$ . Число специализирующих связей в последовательности не ограничено. При этом имена связей, предшествующие последнему, заключены в круглые скобки, а перед открывающей скобкой каждой пары скобок — сдвоенное двоеточие.  $\square$  *Обобщение S-типа Z* — порождение его *S-надтипа Z[#rule]* путем ослабления (здесь # — символ ослабления) связи  $rule$  из семейства связей, соответствующей типу  $Z$ . Исключение связи считаем ее предельным ослаблением.  $\square$

$\square$  **Определение S-системы понятий** — описание ее S-модели, сопровождаемое указанием области применимости.

$\langle$ S-система  $S_c$  понятий $\rangle \approx \langle$ множество  $S^{Sc}$  понятий, на котором определяется изучаемый S-объект $\rangle$ ,  $\langle$ семейство  $rel(S^{Sc})$  связей, заданных на  $S^{Sc}$  $\rangle$ .

Описание представлено в форме S-сообщения, рассчитанного на интерпретацию ИТ-исследователями и разработчиками; представление, сохранение, распространение, накопление и поиск в S-среде.  $\square$

$\diamond$  *Определение S-системы понятий* должно удовлетворять необходимым требованиям конструктивности: представление в виде пары  $\langle$ S-модель системы понятий $\rangle$ ,  $\langle$ определение области S-применимости $\rangle$ ; в S-систему понятий, считающуюся определенной, не должны входить понятия, не имеющие определений (и при этом не относящиеся к понятиям-аксиомам).  $\diamond$

$\square$  *Определение области S-применимости* — описание типов: *корреспондента* (кому адресовано определение); *цели*, в процессе достижения которой определение имеет смысл (классы задач, при изучении которых определение может быть полезно); *стадии*, на которой целесообразно использовать определение (концепция, методология решения и т. д.).  $\square$

*Область S-применимости* может принадлежать совокупности областей, в которых исследуются природные и/или изобретаемые объекты.

$\circ$   $\langle$ S-система  $tr$  понятий *треугольник* $\rangle \approx \langle$ множество  $S^{tr}$  понятий $\rangle$ ,  $\langle$ семейство связей  $rel(S^{tr})$ , заданных на  $S^{tr}$  $\rangle$ . В  $tr$  элементами множества  $S^{tr}$  служат стороны треугольника  $(a, b, c)$ , периметр  $p$  и др. Семейство связей  $rel(S^{tr})$  включает связь  $p = a + b + c$  и др.  $\circ$

Чтобы определить новое понятие (не относящееся к понятиям-аксиомам), необходимо указать его связь с некоторой уже существующей S-системой понятий.

$\circ$  S-система  $tr^{\pi/2}$  понятий *прямоугольный треугольник* может быть определена как *специализация S-типа  $tr : tr^{\pi/2} \approx tr[:(angle = \pi/2)]$* .  $\circ$

$\circ$   $message[:(interface = h \rightarrow m)::means = touch]$  — это специализация S-типа  $message[:(interface = h \rightarrow m)]$ , определяющего множество S-сообщений, соответствующих интерфейсу *человек – S-машина*, по связи  $means = touch$ , выделяющей множество S-сообщений, вводимых в S-машину посредством прикосновений ( $\circ$  к клавишам клавиатуры или сенсорному экрану  $\circ$ ).  $\circ$

### 3 Заключение

Представлена первая из четырех частей теории S-символов, содержащая описание ее концептуальных оснований. Теория S-символов, представляющая собой расширенное обобщение теории S-моделирования [2], рассматривается как часть методологического обеспечения разработок систем искусственного интеллекта (○ систем знаний, систем автоматизации программирования и др. ○) в цифровой среде.

### Литература

1. Newell A., Simon H. A. Computer science as empirical inquiry: Symbols and search // Commun. ACM, 1976. Vol. 19. No. 3. P. 113–126. doi: 10.1145/360018.360022.
2. Plyin V. D. Symbolic modeling (s-modeling): An introduction to theory // Artificial intelligence trends in systems / Ed. R. Silhavy. — Lecture notes in networks and systems ser. — Cham: Springer, 2022. Vol. 502. P. 585–591. doi: 10.1007/978-3-031-09076-9\_54.
3. Ильин В. Д. Система порождения программ. — М.: Наука, 1989. 264 с.
4. Ильин В. Д., Соколов И. А. Символьная модель системы знаний информатики в человеко-автоматной среде // Информатика и её применения, 2007. Т. 1. Вып. 1. С. 66–78.
5. Ильин В. Д. Символьное моделирование // Большая российская энциклопедия (электронная версия). [http://old.bigenc.ru/technology\\_and\\_technique/text/4010980](http://old.bigenc.ru/technology_and_technique/text/4010980).
6. Rojek I., Mikolajewski D., Dostatni E. Digital twins in product lifecycle for sustainability in manufacturing and maintenance // Appl. Sci. — Basel, 2020. Vol. 11, No. 1. 31 p. doi: 10.3390/app11010031.
7. Semeraro C., Lezoche M., Panetto H., Dassisti M. Digital twin paradigm: A systematic literature review // Comput. Ind., 2021. Vol. 130. Art. 103469. 23 p. doi: 10.1016/j.compind.2021.103469.
8. Nguyen H., Trestian R., To D., Tatipamula M. Digital twin for 5G and beyond // IEEE Commun. Mag., 2021. Vol. 59. No. 2. P. 10–15. doi: 10.1109/MCOM.001.2000343.
9. Jia P., Wang X., Shen X. Digital-twin-enabled intelligent distributed clock synchronization in industrial IoT systems // IEEE Internet Things, 2021. Vol. 8. No. 6. P. 4548–4559. doi: 10.1109/JIOT.2020.3029131.

Поступила в редакцию 09.02.23

## THEORY OF S-SYMBOLS: CONCEPTUAL FOUNDATIONS

V. D. Ilyin

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

**Abstract:** The proposed 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 tasks and program design, etc.). The S-environment based on interconnected S-systems (symbols, codes, and signals) serves as the infrastructural basis for the implementation of information technologies for various purposes. The article presents the first part (out of four) of the description of the theory. The rationale for expediency and basic concepts (S-symbol, S-code, S-signal, etc.) are given. The kinds and types of S-(symbols, codes, and signals) are defined. Equivalence, order, and membership relations are introduced, defined on S-systems (symbols, codes, and signals). Definitions are accompanied by examples.

**Keywords:** theory of S-symbols; S-symbol; S-code; S-signal; S-environment; artificial intelligence; information technology

**DOI:** 10.14357/08696527230112

### References

1. Newell, A., and H. A. Simon. 1976. Computer science as empirical inquiry: Symbols and search. *Commun. ACM* 19(3):113–126. doi: 10.1145/360018.360022.
2. Ilyin, V. D. 2022. Symbolic modeling (s-modeling): An introduction to theory. *Artificial intelligence trends in systems*. Ed. R. Silhavy. Lecture notes in networks and systems ser. Cham: Springer. 502:585–591. doi: 10.1007/978-3-031-09076-9\_54.
3. Ilyin, V. D. 1989. *Sistema porozhdeniya programm* [The system of program generating]. Moscow: Nauka. 264 p.
4. Ilyin, V. D., and I. A. Sokolov. 2007. Simvol'naya model' sistemy znaniy informatiki v cheloveko-avtomatnoy srede [The symbol model of informatics knowledge system in human-automaton environment]. *Informatika i ee Primeneniya — Inform. Appl.* 1(1):66–78.
5. Ilyin, V. D. Simvol'noe modelirovanie [Symbolic modeling]. *Bol'shaya russiyskaya entsiklopediya (elektronnaya versiya)* [Great Russian encyclopedia (electronic version)]. Available at: [https://old.bigenc.ru/technology\\_and\\_technique/text/4010980](https://old.bigenc.ru/technology_and_technique/text/4010980) (accessed April 13, 2023).
6. Rojek, I., D. Mikołajewski, and E. Dostatni. 2020. Digital twins in product lifecycle for sustainability in manufacturing and maintenance. *Appl. Sci. — Basel* 11(1). 31 p. doi: 10.3390/app11010031.

7. Semeraro, C., M. Lezoche, H. Panetto, and M. Dassisti. 2021. Digital twin paradigm: A systematic literature review. *Comput. Ind.* 130:103469. 23 p. doi: 10.1016/j.compind.2021.103469.
8. Nguyen, H., R. Trestian, D. To, and M. Tatipamula. 2021. Digital twin for 5G and beyond. *IEEE Commun. Mag.* 59(2):10–15. doi: 10.1109/MCOM.001.2000343.
9. Jia, P., X. Wang, and X. Shen. 2021. Digital-twin-enabled intelligent distributed clock synchronization in industrial IoT systems. *IEEE Internet Things* 8(6):4548–4559. doi: 10.1109/JIOT.2020.3029131.

*Received February 9, 2023*

## **Contributor**

**Ilyin Vladimir D.** (b. 1937) — Doctor of Science in technology, professor, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; vdilyin@yandex.ru

## МЕХАНИЗМ ФОРМИРОВАНИЯ ГИПОТЕЗ В ТЕХНОЛОГИИ ПОДДЕРЖКИ КОНКРЕТНО-ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЙ

*И. М. Адамович<sup>1</sup>, О. И. Волков<sup>2</sup>*

**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. Статья посвящена дальнейшему развитию технологии за счет интеграции в нее механизма, позволяющего автоматически определить потенциально перспективные направления исследования. Предлагаемый подход заключается в автоматическом заполнении информационных лагун в наборе фактов, описывающих объект исследования, на базе неполной индукции. Проведен анализ базы для индуктивного обобщения, показаны способы ее формирования. Обоснована возможность использования для этой цели процедуры импутации данных, используемой в задачах анализа данных и машинного обучения. Проанализированы методы импутации данных с точки зрения особенностей технологии и специфики конкретно-исторического исследования. Анализ показал целесообразность построения механизма автоматического формирования гипотезы с опорой на такой метод импутации данных, как метод деревьев классификации на базе алгоритма CHAID (Chi Squared Automatic Interaction Detection).

**Ключевые слова:** конкретно-историческое исследование; распределенная технология; формирование гипотез; информационная лагуна; импутация данных

**DOI:** 10.14357/08696527230113

### 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

---

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Adam@amsd.com

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Volkov@amsd.com

в этой технологии организованы в форме семантической сети. Узлы сети представляют собой именованные универсальные классы объектов. Факты задаются значениями экземпляров классов и связями между ними. Связи наследуются из сети классов [4].

Данные, описывающие историко-биографические факты, обладают свойством фрагментарности и противоречивости. Получение этих фактов — необходимая и наиболее трудоемкая стадия биографического исследования, связанная с работой в архивах и просмотром огромного числа документов. Дальнейшая работа состоит в сопоставлении фактов между собой и с нормальными (наборами правил, регламентирующих зависимости между фактами) с целью взаимоувязывания и разрешения противоречий [5].

Технология ПКИИ содержит развитые средства автоматизации задач, решаемых пользователем на всех стадиях исследования, включая автоматическую обработку исторических источников [6]. Имеются средства выявления противоречий в данных [7], но такое свойство исторических фактов, как фрагментарность, требует от пользователя трудоемкого поиска недостающей информации и ручного заполнения информационных лакун. Как показано в [8], при выполнении этого процесса исследователю приходится многократно переформулировать информационный запрос, пока результаты поиска его не удовлетворят, т. е. осуществлять итерационный поиск. Анализ результатов, найденных на предыдущем шаге итерации, уточняет направление поиска на последующем шаге. Таких уточнений на одном шаге может быть несколько, поэтому процесс заполнения информационных лакун является не только итерационным, но и ветвящимся процессом. Разработка подхода, облегчающего этот процесс, ставится целью данной статьи.

## **2 Подход к сужению диапазона направлений поиска**

Как показано выше, процесс заполнения информационных лакун является итерационным и ветвящимся и состоит в чередовании процедур:

- (1) определения направлений поиска (формулирования информационного запроса);
- (2) осуществления поиска в выбранном направлении (выполнения информационного запроса).

Технология ПКИИ содержит средства автоматизации и поддержки процедуры осуществления поиска [8], позволяющие отчасти уменьшить трудозатраты исследователя, хотя полностью автоматизировать этот процесс принципиально невозможно. Но выбор направления поиска остается полностью ручной неформальной процедурой, зависящей исключительно от опыта и квалификации исследователя. Неверный выбор направления поиска неизбежно приводит к значительному росту его объемов и к заметному увеличению сроков всего исследования. Поэтому механизмы, позволяющие предложить пользователю перспективные направления поиска, крайне востребованы в технологии ПКИИ.



Предлагаемый подход заключается в автоматическом формировании гипотезы о значении узла семантической сети, соответствующего лакуне в наборе фактов, описывающих объект исследования, на базе неполной индукции. Ценность такого решения для технологии ПКИИ заключается в важности обоснованной гипотезы как метода развития любого научного знания. Особенно это относится к историческому знанию, так как любое знание о прошлом несет в себе долю гипотетичности, поскольку историк не может наблюдать непосредственно исторические события как главный объект своего изучения [9]. Целью предлагаемого подхода ставится не получение новой достоверной информации об объекте исследования, поскольку неполная индукция позволяет делать только вероятностные заключения, а создание обоснованной гипотезы, которая задаст новое направление поиска.

### 3 База для индуктивного обобщения

Повышение вероятности выводов в неполных индуктивных умозаклчениях зависит от соблюдения следующих условий [10]:

- если для индуктивного обобщения взято возможно большее число случаев, относящихся к данному обобщаемому классу явлений;
- когда предметы, знания о которых индуктивно обобщаются, обладают внутренней объективной связью между собой.

В технологии ПКИИ объектами обобщения становятся историко-биографические факты, которые, как сказано выше, задаются значениями экземпляров классов и связями между ними. Связи наследуются из сети классов. Выполнение требования наличия внутренней объективной связи между этими объектами обеспечивается соотносением всех экземпляров и связей, составляющих факты, с одними и теми же универсальными классами и связями между ними. Иными словами, все факты из группы обобщаемых объектов наследуются от одной и той же структуры связанных универсальных классов из подсети понятий семантической сети технологии [2].

Объем группы обобщаемых однотипных объектов (фактов), достаточный для обоснованного вывода, не всегда может быть обеспечен в технологии ПКИИ, поскольку, как показано выше, получение этих фактов — необходимая и наиболее трудоемкая стадия биографического исследования, связанная с работой в архивах и просмотром огромного числа документов. Факты, собираемые вручную одним пользователем технологии в рамках одного конкретно-исторического исследования, как правило, оказываются по своей природе единичными и разрозненными и не могут составить достаточную по объему однородную выборку. Но поскольку технология ПКИИ является распределенной и содержит средства автоматической обработки исторических источников, ее база данных вполне может содержать множественные однородные факты в количестве, вполне достаточном для обоснованного вывода в неполных индуктивных умозаклчениях. Источниками множественных фактов в технологии ПКИИ могут служить:

- (1) *автоматические обработки нарративов*. Технология ПККИИ имеет в своем составе систему Т-парсер [6], предназначенную для автоматического извлечения фактов из текстов историко-биографической направленности на естественном языке повествовательного характера (нарративных источников). Т-парсер ориентирован на многопользовательский распределенный режим работы, что предусматривает одновременное независимое создание правил выделения фактов множеством не связанных между собой исследователей;
- (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 может служить источником множественных однотипных фактов. Обработка может проводиться в рамках одного исследования с целью дальнейшего выявления родственных связей между упомянутыми в реестре персонами как полностью вручную, так и с использованием интегрированных в технологию ПККИИ средств автоматизации исследования:
  - описанной выше системы Т-парсер;
  - инструментального комплекса  $\pi$ -Фактограф [3], ориентированного на поддержку самого исследовательского процесса и решающего задачи фиксации фактов и предположений, поддержки сочетания формализованного и неформализованного представления знаний, отслеживания процессов накопления данных, выявления дефектов, выдвижения и опровержения гипотез;
- (3) *разрозненные одиночные номинативные источники*. Поскольку технология ПККИИ была построена на принципах сотворчества и краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий), ее данные, организованные в форме семантической сети, пополняются независимо разными 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.

#### 4 Импутация данных

Лакуны в данных — известная проблема в задачах анализа данных и машинного обучения. Удаление из набора данных всех неполных наблюдений, как правило, оказывается нежелательным способом решения этой проблемы, поскольку при этом теряется доступ к полезной информации, содержащейся

в непропущенных значениях наблюдений. Альтернативный подход — восстановление пропущенных данных на основании остальных данных в наборе. Эта процедура называется импутацией данных, и для нее разработан ряд методов [11]. Эти методы в принципе могут быть применены для задач автоматического формирования гипотезы о значении узла семантической сети технологии ПКИИ, соответствующего лакуне в наборе фактов, описывающих объект исследования. Для этого группа обобщаемых однотипных объектов (фактов), представленных в форме значений экземпляров классов и их связей, унаследованных от одной и той же структуры связанных универсальных классов из подсети понятий семантической сети технологии, должна быть преобразована в принятую для задач анализа данных табличную форму, где строки и столбцы определены наблюдениями и показателями изучаемой области. Наблюдения соответствуют однородным фактам из анализируемой группы. Показатели соответствуют значениям экземпляров, для которых должен быть определен порядок формирования показателей таблицы. Этот порядок целесообразно строить на базе известного алгоритма обхода дерева или графа, называемого «поиском в глубину» (depth-first search, DFS). Алгоритм DFS — это рекурсивный алгоритм обхода графа. Он принимает на вход некоторую начальную вершину графа и рекурсивно запускает себя для всех еще не посещенных соседей данной вершины. Поскольку, как показано в [5], подсеть семантической сети технологии ПКИИ, соответствующая одному историко-биографическому факту, представляет собой иерархическую структуру, в качестве начальной вершины должен быть выбран корневой узел иерархии.

Помимо очевидных достоинств, импутирование как способ решения проблемы недостающей информации имеет два принципиальных недостатка, которые нельзя не учитывать при анализе данных [12]:

- (1) как правило, параметры для алгоритма заполнения пропусков вычисляются по присутствующим данным, что вносит зависимость между наблюдениями;
- (2) распределение данных после заполнения будет отличаться от истинного, даже если пренебречь зависимостью, указанной выше.

Анализ подобных полных данных стандартными методами неправомерен и приводит к таким недостаткам, как несостоятельность и смещенность оценок параметров. Но для целей конкретно-исторического исследования в рамках технологии ПКИИ данный подход вполне применим. Поскольку целью применения импутации данных ставится не формирование оценок, принимаемых за истину, а построение гипотезы, подлежащей дальнейшей проверке всеми методами конкретно-исторического исследования, приведенные выше недостатки подхода оказываются несущественными.

## 5 Методы импутации данных

В [12] приведены основные методы импутации данных. К глобальным методам (методам заполнения лакун с использованием данных всей выборки) относятся:

- заполнение средними;
- методы заполнения с подбором;
- заполнение по регрессии;
- метод сплайн-интерполяции;
- методы многократного заполнения;
- МП-оценивание (EM-алгоритм);
- использование методов факторного анализа;
- использование методов кластерного анализа.

Локальные алгоритмы опираются на гипотезу локальной компактности, которая состоит в утверждении, что для предсказания пропущенного элемента  $b_{ij}$  нужно использовать не всю таблицу, а лишь ее «компетентную» часть, состоящую из элементов строк, похожих на строку  $i$ , и элементов столбцов, похожих на столбец  $j$ . Остальные строки и столбцы для данного элемента неинформативны [13]. К таким алгоритмам относятся:

- алгоритмы семейства ZET;
- алгоритмы семейства WANGA.

Все эти методы и алгоритмы имеют свои достоинства и недостатки, но их общее свойство заключается в том, что они относятся к числовым данным, в то время как наиболее значимой задачей для технологии ПКИИ остается построение гипотез относительно значений качественных, категориальных данных. Предсказание числовых значений каких-то атрибутов тех или иных объектов представляется слабоприменимым в конкретно-историческом исследовании, на поддержку которого ориентирована технология ПКИИ, поскольку большинство изучаемых в рамках исследования параметров очень вариативны. Предсказания (гипотезы относительно значений) могут быть осмысленны только при усреднении значений параметра для множества объектов, что соответствует теоретическому историческому исследованию, для которого характерен высокий уровень обобщения, а не конкретно-историческому исследованию, на первый план в котором выдвигаются задачи описательного характера [14].

В [15] приведен сравнительный анализ методов импутации, которые могут быть применены для восстановления пропущенных значений в массивах категориальных переменных.

1. *Импутация с использованием меры средней тенденции.* Сущность метода заключается в том, что пропущенные значения заполняются подходящими мерами средней тенденции, рассчитанными по набору комплектных записей. В частности, для номинальной переменной на место пропуска подставляется мода, для порядковых — мода или медиана.
2. *Импутация с использованием деревьев классификации.* На сегодняшний день разработаны несколько алгоритмов построения дерева. Для импутации категориальных переменных используется алгоритм CHAID [16]. Этот

алгоритм обладает рядом существенных преимуществ: он универсален и обрабатывает переменные по любым шкалам; не накладывает ограничения на закон распределения величин; результаты, полученные в ходе применения метода, легко интерпретируются исследователем.

3. *Импутация с использованием нейронных сетей.* Все модели искусственных нейронных сетей (ИНС) в процессе своего построения требуют обучения (с учителем, без учителя или смешанное обучение) и настройку весовых коэффициентов. Недостаток метода — значительные временные затраты на построение и анализ сетей. Для решения задачи импутации выбирается ИНС прямого распространения или многослойный перцептрон.

Анализ эффективности методов, проведенный на выборках с дихотомическими, номинальными и порядковыми переменными, содержащими от 5% до 90% пропущенных значений, показал, что метод деревьев классификации на базе алгоритма CHAID позволил получить наилучшие результаты в 90% исследуемых случаях. Восстановление пропусков этим методом позволило получить наименьшую долю неверно импутированных значений и наименьшую среднюю абсолютную погрешность.

## 6 Выводы

Интеграция в технологию ПКИИ механизма автоматического формирования гипотезы о значении узла семантической сети, соответствующего лакуне в наборе фактов, описывающих объект исследования, на базе неполной индукции может существенно упростить и повысить эффективность такой ручной неформальной процедуры, зависящей исключительно от опыта и квалификации исследователя, как выбор направления поиска при проведении конкретно-исторического исследования.

Анализ показал целесообразность построения механизма автоматического формирования гипотезы с опорой на такой метод импутации данных, как метод деревьев классификации на базе алгоритма CHAID, поскольку он позволяет эффективно восстанавливать пропущенные значения в массивах категориальных переменных, что актуально для технологии ПКИИ.

Актуальность дальнейшего развития в рассмотренном направлении 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. *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.
6. *Адамович И. М., Волков О. И.* Система извлечения биографических фактов из текстов исторической направленности // Системы и средства информатики, 2015. Т. 25. № 3. С. 235–250. doi: 10.14357/08696527150315.
7. *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.
8. *Адамович И. М., Волков О. И.* Средства поддержки интернет-поиска при проведении биографических исследований // Системы и средства информатики, 2014. Т. 24. № 2. С. 178–192. doi: 10.14357/08696527140212.
9. *Бочаров А. В.* Алгоритмы использования основных научных методов в конкретно-историческом исследовании. — Томск: ТГУ, 2007. 140 с.
10. *Бесхлебный Е. И.* Логика для юристов. — М.: Юстиция, 2021. 248 с.
11. *Фомина Е. Е.* Обзор методов и программного обеспечения для восстановления пропущенных значений и массивах социологических данных // Гуманитарный вестник, 2019. № 4(78). 12 с. doi: 10.18698/2306-8477-2019-4-611.
12. *Абраменкова И. В., Круглов В. В.* Методы восстановления пропусков и массивах данных // Программные продукты и системы, 2005. № 2. С. 18–22.
13. *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний. — Новосибирск: Изд-во Ин-та математики, 1999. 270 с.
14. *Ипполитов Г. М., Репинецкая Ю. С.* Историческое исследование: логика, стратегия, принципы, методы. — Самара: СГСПУ, 2019. 214 с.
15. *Фомина Е. Е.* Сравнительный анализ методов импутации категориальных переменных в массивах с результатами социологических опросов // Вестник ПНИПУ. Социально-экономические науки, 2021. № 1. С. 83–96.
16. *Жучкова С. В., Ротмистров А. Н.* Возможность работы с пропущенными данными при использовании CHAID: результаты статистического эксперимента // Социология: методология, методы, математическое моделирование, 2018. № 46. С. 85–122.

*Поступила в редакцию 27.11.22*

## HYPOTHESIS FORMATION MECHANISM 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 article is devoted to the further development of the technology by integrating into it a mechanism that automatically identifies potentially promising areas of research. The proposed approach is to automatically fill in information gaps in a set of facts describing the object of research on the basis of incomplete induction. The analysis of the base for inductive generalization is carried out and the ways of its formation are shown. The possibility of using the data imputation procedure usually used in data analysis and machine learning tasks for this purpose is substantiated. The methods of data imputation are analyzed in the connection with the features of technology and the specifics of concrete historical research. The analysis showed the expediency of the mechanism for automatic hypothesis formation constructing through such method of data imputation as the method of classification trees based on the CHAID (Chi Squared Automatic Interaction Detection) algorithm.

**Keywords:** concrete historical investigation; distributed technology; formation of hypotheses; information gap; data imputation

**DOI:** 10.14357/08696527230113

### References

1. Gribach, S.V. 2010. Issledovanie semeynykh krizisov posredstvom psikholingvisticheskogo eksperimenta [The study of family crises through a psycholinguistic experiment]. *Sborniki konferentsiy NIT's Sotsiosfera* [Conference Proceedings NIC Sociosfera] 6:45–54.
2. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelenogo 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. 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.
  6. Adamovich, I. M., and O. I. Volkov. 2015. Sistema izvlecheniya biograficheskikh faktov iz tekstov istoricheskoy napravlenosti [The system of facts extraction from historical texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):235–250. doi: 10.14357/08696527150315.
  7. Adamovich, I. M., and O. I. Volkov. 2020. Avtomatizirovannyi poisk protivorechii 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.
  8. Adamovich, I. M., and O. I. Volkov. 2014. Sredstva podderzhki internet-poiska pri provedenii biograficheskikh issledovaniy [The technology of internet search as a part of biographic investigation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(2):178–192. doi: 10.14357/08696527140212.
  9. Bocharov, A. V. 2007. *Algoritmy ispol'zovaniya osnovnykh nauchnykh metodov v konkretno-istoricheskoy issledovaniy* [Algorithms of basic scientific methods using in concrete historical research]. Tomsk: Publishing House of the Tomsk State University. 140 p.
  10. Beskhlebnyy, E. I. 2021. *Logika dlya yuristov* [Logic for lawyers]. Moscow: Yustitsiya. 248 p.
  11. 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]. *Humanities Bulletin BMSTU* 4(78). 12 p. doi: 10.18698/2306-8477-2019-4-611.
  12. Abramenkova, I. V., and V. V. Kruglov. 2005. Metody vosstanovleniya propuskov v massivakh dannykh [Methods for restoring data gaps in data arrays]. *Programmnye produkty i sistemy* [Software & Systems] 2:18–22.
  13. Zagoruyko, N. G. 1999. *Prikladnye metody analiza dannykh i znaniy* [Applied methods of data and knowledge analysis]. Novosibirsk: Publishing House of Institute of Mathematics. 270 p.
  14. Ippolitov, G. M., and Yu. S. Repinetskaya. 2019. *Istoricheskoe issledovanie: logika, strategiya, printsipy, metody* [Historical research: Logic, strategy, principles, and methods]. Samara: SGSPU. 214 p.
  15. Fomina, E. E. 2021. Sravnitel'nyy analiz metodov imputatsii kategorial'nykh peryemnykh v massivakh s rezul'tatami sotsiologicheskikh oprosov [Comparative analysis of the methods of categorical variables imputation in arrays for sociological surveys]. *Vestnik PNIPU. Sotsial'no-ekonomicheskie nauki* [PNRPU Sociology and Economics Bulletin] 1:83–96. doi: 10.15593/2224-9354/2021.1.7.
  16. Zhuchkova, S. V., and A. N. Rotmistrov. 2018. Vozmozhnost' raboty s propushchennymi dannymi pri ispol'zovanii CHAID: Rezul'taty statisticheskogo eksperimenta [Handling missing data with CHAID: Results of a statistical experiment]. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie* [Sociology: Methodology, Methods, Mathematical Modeling] 46:85–122.

*Received November 27, 2022*



## Contributors

**Adamovich Igor M.** (b. 1934)— Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, 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, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Volkov@amsd.com

## БАЗИСНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИНФОРМАЦИОННАЯ ВОЙНА И МНОГОМЕРНЫЙ ИЕРАРХИЧЕСКИЙ ТЕРРИТОРИАЛЬНЫЙ СУВЕРЕНИТЕТ: ЭТАПЫ ГЛОБАЛЬНО-КОСМИЧЕСКОЙ КОЭВОЛЮЦИИ

С. Н. Гринченко<sup>1</sup>

**Аннотация:** С позиций информатико-кибернетического моделирования процесса развития самоуправляющейся иерархо-сетевой системы Человечества рассмотрен параллелизм процессов глобально-космической коэволюции базисных информационных технологий (БИТ), «многомерного иерархического территориального суверенитета» (МИТС) и информационной войны (ИВ). Периоду формирования субъектами: (1) «дворов»/семей *Homoidea* БИТ сигнальных поз/звуков/движений соответствует развитие «пред-пред-суверенитета» сообществ этих территорий (семейной особи) и «пред-пред-ИВ пред-пред-обмана противника»; (2) «поселений» *Homo erectus* БИТ мимика/жесты соответствует развитие «пред-суверенитета» сообществ этих территорий (родовой особи) и «пред-ИВ пред-обмана противника»; (3) «округ» *Homo sapiens-1* БИТ речи/языка соответствует развитие «языкового яруса» МИТС соответствующих «протоцивилизаций» и «ИВ обмана противника и развала его “округ”»; (4) «сверхрайонов» *Homo sapiens-2* БИТ письменности/чтения соответствует развитие «культурно-государственного яруса» МИТС локальных цивилизаций и ИВ по развалу «сверхрайонов» противника; (5) «сверхстран» *Homo sapiens-3* БИТ тиражирования текстов/книгопечатания соответствует развитие «экономического яруса» МИТС субконтинентальных цивилизаций и ИВ по развалу «сверхстран» противника; (6) создаваемой Планетарной Цивилизации *Homo sapiens-4* БИТ локальных компьютеров соответствует развитие ее «высокотехнологического яруса» МИТС и ИВ по развалу образовываемых противником ее элементов; (7) создаваемой Цивилизации Околоземного Космоса *Homo sapiens-5* БИТ телекоммуникаций/сетей соответствует развитие ее «информационного яруса» МИТС и ИВ по развалу образовываемых противником ее элементов; (8) создаваемой Цивилизации Промежуточного Космоса *Homo sapiens-6* перспективной нано-БИТ соответствует развитие ее «лично-космического яруса» МИТС и ИВ по развалу образовываемых противником ее элементов и т. д. Отмечается, что каждый новый этап этого коэволюционного процесса не отменяет результатов предыдущего, а дополняет и усложняет их.

**Ключевые слова:** базисные информационные технологии; информационное оружие; информационная безопасность; многомерный иерархический территориальный суверенитет; системная коэволюция; информатико-кибернети-

---

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sgrinchenko@ipiran.ru

ческая модель; самоуправляющаяся иерархо-сетевая система Человечества; принцип системной кумуляции; принцип системной согласованности

**DOI:** 10.14357/08696527230114

*Информационные войны* принято рассматривать как противоборство сторон посредством распространения *специально подготовленной информации* и противодействия аналогичному внешнему воздействию на себя. Очевидно, что ИВ между различными государствами и общественными структурами сегодня происходят практически перманентно, меняется лишь в той или иной степени их накал. Внимательный анализ показывает, что предшественники ИВ встречались и в историческом прошлом — пусть и менее ярко выраженные. И тогда возникает вопрос: как они развивались и когда стали столь опасными?

Противодействие ИВ обеспечивается посредством фактора *информационной безопасности* — в ячейках общества от семьи до Человечества как единого системного целого. Типичные размеры подобных ячеек, увеличивающиеся в процессе глобальной эволюции, определяются в рамках авторской информатико-кибернетической модели (ИКМ) [1–4]. Критическими временными точками такой системной эволюции служат *системные перевороты* в истории Человечества, маркерами которых выступают *информационные перевороты*, т. е. старты новых БИТ, с модельными расчетными датировками<sup>1</sup> ~ 28 230–1860–123–8,1 тыс. лет назад–1446–1946–1979–1981–1981 гг.–...

В свою очередь, согласно ИКМ, структура феномена «территориальный суверенитет» общественно-государственных образований в системе Человечества является исторически многомерной, включающей 6 основных иерархических компонент: языковую, культурно-государственную, экономическую, высокотехнологическую, информационную и (перспективную) личностно-космическую [6, 7]. Таким образом, она образует «многомерный иерархический территориальный суверенитет».

Отметим, что важной особенностью ИКМ является *принцип системной кумуляции*, согласно которому возникновение в иерархических природных системах новых подсистем не означает элиминации ранее возникших, а также *принцип системной согласованности*, гласящий, что возникновение в ходе эволюции системы Человечества новых подсистем сопровождается усложнениями в структуре и характере приспособительного поведения ранее возникших [3].

Сопоставляя процессы эволюционного развития БИТ, ИВ и МИТС на базе системного модельного подхода в рамках ИКМ, приходим к представлению об их *глобально-космической коэволюции* (см. таблицу).

Действительно, параллелизм хода усложнения и расширения ареала существования/применения данных сущностей/поведений в историческом процессе

<sup>1</sup>Последовательности временных и пространственных количественных параметров ИКМ базируются на геометрической прогрессии со знаменателем  $e^e = 15,15426 \dots$ , выявленной А. В. Жирмунским и В. И. Кузьминым при исследовании критических уровней в развитии биологических систем [5].

Сопоставление этапов глобально-космической коэволюции БИТ, форм МИТС и ярусов пространственной иерархии ИВ в самоуправляющейся иерархо-сетевой системе Человечества

№	БИТ. Субъект	Время старта, кульминации	Ареал (радиус эквивалентного круга/шара)	МИТС	Пространственная иерархия ИВ
1	Сигнальные позы/звуки/движения. <i>Homoidea</i>	~ 28,2, ~ 9,26 млн лет назад	~ 4,2–64 м, «двор»/семья	Семейная особость — территориальный «пред-пред-суверенитет» «пред-пред-цивилизаций»	«Пред-пред-обман» противника и «пред-пред-ИВ» по развалу его «дворов»/семей
2	Мимика/жесты. <i>Homo erectus</i>	~ 1,86, ~ 0,612 млн лет назад	~ 64 м–1 км, «поселение»	Родовая особость — территориальный «пред-суверенитет» «пред-цивилизаций»	+ «Пред-обман» противника и «пред-ИВ» по развалу его «поселений»
3	Речь/язык. <i>Homo sapiens-1</i>	~ 123, ~ 40,3 тыс. лет назад	~ 1–15 км, «округа»	«Языковой» (ЯТС) — «протоцивилизаций»	+ Обман противника и ИВ по развалу его «округ»
4	Письменность/чтение. <i>Homo sapiens-2</i>	~ 8,1, ~ 2,7 тыс. лет назад	~ 15–223 км, «сверхрайон» (национальное государство)	«Культурно-государственный» (КГТС) + ЯТС — локальных цивилизаций	+ ИВ по развалу национальных «сверхрайонов» противника — локальных цивилизаций
5	Тиражирование текстов/книгопечатание. <i>Homo sapiens-3</i>	~ 1446, ~ 1806 гг.	~ 223–3370 км, «сверхстрана» (наднациональная империя, держава и т. д.)	«Экономический» (ЭТС) + КГТС + ЯТС — субконтинентальных цивилизаций	+ ИВ по развалу наднациональных «сверхстран» противника — субконтинентальных цивилизаций
6	Локальный компьютер. <i>Homo sapiens-4</i>	~ 1946, ~ 1970 гг.	~ 3,37–51 тыс. км, планета Земля	«Высокотехнологический» (ВТС) + ЭТС + КГТС + ЯТС — Планетарной Цивилизации	+ ИВ по развалу формируемых противником элементов Планетарной Цивилизации
7	Телекоммуникации/сети. <i>Homo sapiens-5</i>	~ 1979, ~ 2003 гг.	~ 51–773 тыс. км, Околосредней Космос (пространственный объем)	«Информационный» (ИТС) + ВТС + ЭТС + КГТС + ЯТС — Цивилизации Околосредней Космоса	+ ИВ по развалу формируемых противником элементов Цивилизации Околосредней Космоса
8	Нано-ИТ. <i>Homo sapiens-6</i>	~ 1981, ~ 2341 гг.	~ 0,773–11,7 млн км, Промежуточный Космос (пространственный объем)	«Личностно-космический» (ЛКТС) + ИТС + ВТС + ЭТС + КГТС + ЯТС — Цивилизации Промежуточного Космоса	+ ИВ по развалу формируемых противником элементов Цивилизации Промежуточного Космоса
...	...	...	...	...	...

развития Человечества трудно не отметить, он прослеживается с древнейших времен.

Так, еще при развитии доречевых БИТ формирующиеся сообщества «пред-пред-человека» и «пред-человека» как-то взаимодействовали между собой, и естественно предположить, что при отстраивании компонент «пред-МИТС» использовались и соответствующие формы «пред-пред-ИВ» и «пред-ИВ».

На следующем этапе глобальной эволюции (~ 123–8,1 тыс. лет назад) формирующиеся на базе БИТ *речи/языка* сообщества Homo sapiens-1 (на территориях радиусом эквивалентного круга до 15 км — «округ», «протоцивилизаций»), отстаивая свою — *языковую* — компоненту МИТС, используют обман противника и другие формы ИВ по развалу его «округ». Понятно, что использование лишь БИТ речи/языка ограничивает возможности используемой при этом ИВ.

На следующем этапе глобальной эволюции (~ 8,1 тыс. лет назад — ~ 1446 г.) формирующиеся на базе БИТ *письменности/чтения* сообщества Homo sapiens-2 (на территориях радиусом до 223 км — «сверхрайонов» — национальных государств, локальных цивилизаций), отстаивая свою — *культурно-государственную* — компоненту МИТС, используют ИВ по развалу «сверхрайонов» противника и всех их иерархических составляющих. Следует отметить, что отстаивание своего культурно-государственного суверенитета остается актуальной задачей и в наши дни.

На следующем этапе глобальной эволюции (~ 1446 — ~ 1946 гг.) формирующиеся на базе БИТ *тиражирования текстов/книгопечатания* сообщества Homo sapiens-3 (на территориях радиусом до 3370 км — «сверхстран» — наднациональных государств, субконтинентальных цивилизаций), отстаивая свою — *экономическую* — компоненту МИТС, используют ИВ по развалу «сверхстран» противника и всех их иерархических составляющих. Следует также отметить, что отстаивание своего экономического суверенитета остается более чем актуальной задачей и в наши дни.

На следующем этапе глобальной эволюции (~ 1946 — ~ 1979 гг. и далее) формирующиеся на базе БИТ *локальных компьютеров* сообщества Homo sapiens-4 (на территориях радиусом до 51 тыс. км — на планете Земля в целом, Планетарной Цивилизации), отстаивая свою — *высокотехнологическую* — компоненту МИТС, используют ИВ по развалу формируемых противником элементов Планетарной Цивилизации и всех их иерархических составляющих. Следует также отметить, что отстаивание своего высокотехнологического суверенитета остается актуальнейшей задачей и сегодня.

На следующем этапе глобально-космической эволюции (~ 1979 — ~ 1981 гг. и далее) формирующиеся на базе БИТ *телекоммуникаций/сетей* сообщества Homo sapiens-5 (на территориях радиусом до 773 тыс. км — в Околосемном Космосе), отстаивая свою — *информационную* — компоненту МИТС, используют ИВ по развалу формируемых противником элементов Цивилизации Околосемного Космоса и всех их иерархических составляющих. Очевидно, что отстаивание

своего информационного суверенитета остается сегодня **первоочередной задачей** всех участников цивилизационной гонки.

На следующем этапе глобально-космической эволюции (~ 1981 г. и далее) формирующиеся на базе *нано-БИТ* сообщества *Homo sapiens-6* (на территориях радиусом до 11,7 млн км — в Промежуточном Космосе), отстаивая свою — *лично-космическую* — компоненту *МИТС*, используют *ИВ* по развалу формируемых противником элементов Цивилизации Промежуточного Космоса и всех их иерархических составляющих. Имея в виду, что *нано-БИТ* — это перспективная технология (ее кульминацию ИКМ ожидает около 2341 г. [8, 9]), следует полагать, что отстаивание лично-космического суверенитета в рамках этой будущей Цивилизации станет задачей перспективной.

И так далее.

В целом еще раз [4] приходим к констатации факта наличия коэволюционных связей между различными составляющими единого системного процесса глобально-космической эволюции в системе Человечества, в данном случае форм реализации БИТ, разновидностей происходящих в ней ИВ и конфигураций МИТС составляющих ее общественных и государственных образований.

Таким образом, предлагаемая модель параллелизма в глобальной коэволюции БИТ, МИТС и ИВ в самоуправляющейся иерархо-сетевой системе Человечества, верификация основных положений которой проведена на имеющихся историко-археологических данных [10–12], может быть положена в основу и новых подходов к изучению сущности Человека/Человечества как имманентных составляющих Мироздания и расширения методов практического обеспечения информационной безопасности общественно-государственных структур.

## Литература

1. Гринченко С. Н. Метаэволюция (систем неживой, живой и социально-технологической природы). — М.: ИПИРАН, 2007. 456 с.
2. Гринченко С. Н. Базисные информационные технологии и «промышленные революции»: какова их взаимосвязь? // Системы и средства информатики, 2020. Т. 30. № 3. С. 164–172.
3. Гринченко С. Н. Общение людей и информационные технологии: принципы системной кумуляции и системной согласованности // Мир психологии, 2020. № 3(103). С. 235–245.
4. Гринченко С. Н. О сравнительном генезисе базисных информационных технологий и общественно-экономических формаций Человечества // Системы и средства информатики, 2022. Т. 32. № 3. С. 147–155.
5. Жирмунский А. В., Кузьмин В. И. Критические уровни в процессах развития биологических систем. — М.: Наука, 1982. 179 с.
6. Гринченко С. Н. Феномен «территориального суверенитета» с кибернетических позиций: многомерность, иерархическая структура, глобальная эволюция // Ученый совет, 2022. Т. XIX. № 5(209). С. 323–329.

7. *Гринченко С. Н.* О глобальной эволюции феномена «многомерный иерархический территориальный суверенитет» и территориальных структур самоуправления системы Человечества: кибернетическое представление // Моисеевские чтения: Стратегическое целеполагание, формирование нового мировоззрения и образование. V Общеросс. (национальная) научн. конф. — М.: МосГУ, 2022. Т. I. С. 51–59.
8. *Гринченко С. Н.* Развитие базисных информационных технологий и системная сингулярность: кибернетическая точка зрения // Вестник Восточно-Сибирской открытой академии, 2020. № 39. 10 С.
9. *Grinchenko S. N., Shchapova Y. L.* The deductive approach to Big History’s Singularity // The 21st century singularity and global futures: A Big History perspective / Eds. A. V. Korotayev, D. LePoire. — World-systems evolution and global futures ser. — Cham: Springer, 2020. P. 201–210.
10. *Гринченко С. Н., Щапова Ю. Л.* Информационные технологии в истории Человечества // Информационные технологии, 2013. № S8. С. 1–32.
11. *Щапова Ю. Л., Гринченко С. Н.* Введение в теорию археологической эпохи: числовое моделирование и логарифмические шкалы пространственно-временных координат. — М.: Ист. фак. МГУ им. М. В. Ломоносова; ФИЦ ИУ РАН, 2017. 236 с.
12. *Щапова Ю. Л., Гринченко С. Н., Кокорина Ю. Г.* Информатико-кибернетическое и математическое моделирование археологической эпохи: логико-понятийный аппарат. — М.: ФИЦ ИУ РАН, 2019. 136 с.

*Поступила в редакцию 31.10.22*

---

---

## **BASIC INFORMATION TECHNOLOGIES, INFORMATION WARFARE, AND MULTIDIMENSIONAL HIERARCHICAL TERRITORIAL SOVEREIGNTY: STAGES OF GLOBAL-SPACE COEVOLUTION**

*S. N. Grinchenko*

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

**Abstract:** From the standpoint of informatics-cybernetic modeling of the process of development of a self-controlling hierarchical-network system of Humankind, the parallelism of the processes of global-space coevolution of basic information technologies (BIT), “multidimensional hierarchical territorial sovereignty” (MHTS), and information warfare (IW) is considered. The period of formation by the subjects of: (i) “courtyards”/families of Hominoidea of signal postures/sounds/movements BIT corresponds to the development of “pre-pre-sovereignty” of the communities of these territories (family specificity) and “pre-pre-IW pre-pre-deception of the enemy”; (ii) “settlements” of Homo erectus of mimics/gestures BIT correspond to the development of “pre-sovereignty” of communities of these territories (generic identity) and “pre-IW pre-deception of



the enemy”; (iii) “environments” of Homo sapiens-1 of speech/language BIT corresponds to the development of the MHTS “linguistic tiers” of the corresponding proto-civilizations and “IT of deception of the enemy and the collapse of his “environment”; (iv) “superregions” of Homo sapiens-2 of writing/reading BIT corresponds to the development of the MHTS “cultural-state tiers” of local civilizations and IW for the collapse of the enemy’s “superregions;” (v) “super-countries” of Homo sapiens-3 of text replication/printing BIT corresponds to the development of the MHTS “economic tiers” of subcontinental civilizations and IW to destroy the enemy’s “supercountries”; (vi) the created Planetary Civilization of Homo sapiens-4 of local computers BIT corresponds to the development of its MHTS “high-tech tiers” and IW for the collapse of its elements formed by the enemy; (vii) the created Civilization of the Near-Earth Space of Homo sapiens-5 of telecommunications/networks BIT corresponds to the development of its MHTS “information tiers” and IW for the collapse of its elements formed by the enemy; (viii) the emerging Civilization of the Intermediate Cosmos of Homo sapiens-6 with a promising nano-BIT corresponds to the development of its MHTS “personal-cosmic tiers” and IW for the collapse of its elements formed by the enemy; etc. It is noted that each new stage of this coevolutionary process does not cancel the results of the previous one but complements and complicates them.

**Keywords:** basic information technologies; information weapon; information security; multidimensional hierarchical territorial sovereignty; systemic coevolution; informatics-cybernetic model; self-controlling hierarchical-network system of Humankind; principle of systemic cumulation; principle of system consistency

**DOI:** 10.14357/08696527230114

## References

1. Grinchenko, S.N. 2007. *Metaevolyutsiya (sistem nezhyvoy, zhivoy i sotsial'no-tekhnologicheskoy prirody)* [Meta-evolution (of inanimate, animate, and socio-technological nature systems)]. Moscow: IPIRAN. 456 p.
2. Grinchenko, S.N. 2020. Bazisnye informatsionnye tekhnologii i «promyshlennye revolyutsii»: Kakova ikh vzaimosvyaz'? [Basic information technologies and “industrial revolutions”: What is their relationship?]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(3):164–172.
3. Grinchenko, S.N. 2020. Obshchenie lyudey i informatsionnye tekhnologii: Printsipy sistemnoy kumulyatsii i sistemnoy soglasovannosti [Communication of people and information technology: Principles of systemic cumulation and systemic coordination]. *Mir psikhologii* [The World of Psychology] 3(103):235–245.
4. Grinchenko, S.N. 2022. O sravnitel'nom genezise bazisnykh informatsionnykh tekhnologiy i obshchestvenno-ekonomicheskikh formatsiy Chelovechestva [On the comparative genesis of basic information technologies and socioeconomic formations of the Humankind]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):147–155.
5. Zhirmunskiy, A. V., and V. I. Kuz'min. 1982. *Kriticheskie urovni v protsessakh razvitiya biologicheskikh sistem* [Critical levels in the development of biological systems]. Moscow: Nauka. 179 p.



6. Grinchenko, S. N. 2022. Fenomen “territorial’nogo suvereniteta” s kiberneticheskikh pozitsiy: Mnogomernost’, ierarkhicheskaya struktura, global’naya evolyutsiya [The phenomenon of “territorial sovereignty” from the cybernetic perspective: Multidimensionality, hierarchical structure, global evolution]. *Uchenyy sovet* [Academic Council] 19(5):323–329.
7. Grinchenko, S. N. 2022. O global’noy evolyutsii fenomena “mnogomernyy ierarkhicheskiy territorial’nyy suverenitet” i territorial’nykh struktur samoupravleniya sistemy Chelovechestva: kiberneticheskoe predstavlenie [On the global evolution of the phenomenon of “multidimensional hierarchical territorial sovereignty” and territorial self-controlling structures of the Humanity system: A cybernetic representation]. *Moiseevskie chteniya: Strategicheskoe tselepolaganie, formirovanie novogo mirovozzreniya i obrazovanie: V Obshcheross. (natsional’naya) nauchn. konf.* [Moiseev Readings: Strategic goal-setting, the formation of a new worldview and education. 5th All-Russian (National) Scientific Conference] 1:51–59.
8. Grinchenko, S. N. 2020. Razvitie bazisnykh informatsionnykh tekhnologiy i sistemnaya singulyarnost’: Kiberneticheskaya tochka zreniya [Development of basic information technologies and system singularity: A cybernetic point of view]. *Vestnik Vostochno-Sibirskoy otkrytoy akademii* [Bulletin of the East Siberian Open Academy] 39. 10 p.
9. Grinchenko, S. N., and Y. L. Shchapova. 2020. The deductive approach to Big History’s Singularity. *The 21st century singularity and global futures: A Big History perspective*. Eds. A. V. Korotayev and D. LePoire. World-systems evolution and global futures ser. Cham: Springer. 201–210.
10. Grinchenko, S. N., and Y. L. Shchapova. 2013. Informatsionnye tekhnologii v istorii Chelovechestva [Information technologies in the history of Humankind]. *Informatsionnye tekhnologii* [Information Technologies] S8:1–32.
11. Shchapova, Y. L., and S. N. Grinchenko. 2017. *Vvedenie v teoriyu arkheologicheskoy epokhi: Chislovoe modelirovanie i logarifmicheskie shkaly prostranstvenno-vremennykh koordinat* [Introduction to the theory of the archaeological epoch: Numerical modeling and logarithmic scales of space–time coordinates]. Moscow: Faculty of History MSU, FRC CSC RAS. 236 p.
12. Shchapova, Yu. L., S. N. Grinchenko, and Yu. G. Kokorina. 2019. *Informatiko-kiberneticheskoe i matematicheskoe modelirovanie arkheologicheskoy epokhi: Logikoponyatiynyy apparat* [Informatics-cybernetic and mathematical modeling of the archaeological epoch: Logical-conceptual apparatus]. Moscow: FRC CSC RAS. 136 p.

Received October 31, 2022

## Contributor

**Grinchenko Sergey N.** (b. 1946) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sgrinchenko@ipiran.ru

---

## ОБ АВТОРАХ

---

**Адамович Игорь Михайлович** (р. 1934) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Бутенко Юлия Ивановна** (р. 1987) — кандидат технических наук, доцент кафедры теоретической информатики и компьютерных технологий Московского государственного технического университета им. Н. Э. Баумана

**Волков Олег Игоревич** (р. 1964) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Гринченко Сергей Николаевич** (р. 1946) — доктор технических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Николай Александрович** (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Дурново Александр Адрианович** (р. 1949) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Дьяченко Денис Юрьевич** (р. 1987) — инженер-исследователь Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Дьяченко Юрий Георгиевич** (р. 1958) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Егоров Владимир Борисович** (р. 1948) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Забезжайло Михаил Иванович** (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дород-

ница 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

**Иванов Владимир Алексеевич** (р. 1957) — доктор военных наук, профессор, член-корреспондент Академии криптографии Российской Федерации; профессор МИРЭА — Российского технологического университета

**Ильин Владимир Дмитриевич** (р. 1937) — доктор технических наук, профессор, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Инькова Ольга Юрьевна** (р. 1965) — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; сотрудник Университета Женева

**Ковалёв Иван Александрович** (р. 1996) — аспирант Вологодского государственного университета; научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Коньшев Михаил Юрьевич** (р. 1975) — доктор технических наук, доцент, профессор Департамента информационной безопасности Финансового университета при Правительстве Российской Федерации

**Кривенко Михаил Петрович** (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кружков Михаил Григорьевич** (р. 1975) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Мартюшова Янина Германовна** (р. 1970) — кандидат педагогических наук, доцент кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального исследовательского университета)

**Минеева Татьяна Андреевна** (р. 2002) — студентка бакалавриата Московского авиационного института (национального исследовательского университета)

**Морозов Николай Викторович** (р. 1956) — старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Наумов Андрей Викторович** (р. 1966) — доктор физико-математических наук, профессор кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального исследовательского университета)

**Нуриев Виталий Александрович** (р. 1980) — доктор филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Смирнов Сергей Владиславович** (р. 1976) — кандидат технических наук, доцент МИРЭА — Российского технологического университета

**Соколов Игорь Анатольевич** (р. 1954) — доктор технических наук, академик РАН, директор Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Степченков Юрий Афанасьевич** (р. 1951) — кандидат технических наук, заведующий отделом, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Сучков Александр Павлович** (р. 1954) — доктор технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Тараканов Олег Викторович** (р. 1966) — кандидат технических наук, доцент МИРЭА — Российского технологического университета

**Тараканова Валерия Олеговна** (р. 1997) — сотрудник ФСО России

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Усовик Сергей Викторович** (р. 1982) — кандидат технических наук, старший преподаватель МИРЭА — Российского технологического университета

**Шоргин Сергей Яковлевич** (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики 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<sup>A</sup>T<sub>E</sub>X (.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](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](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 rashoda 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 Pubs. 10 p.

**Описание патента:**

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.

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](mailto: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<sup>A</sup>T<sub>E</sub>X (.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](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](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 slovarya 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](mailto:ssi@frccsc.ru) (to Svetlana Strigina)

[http://www.ipiran.ru/english/journal\\_systems.asp](http://www.ipiran.ru/english/journal_systems.asp)

# SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

**Volume 33 No.1 Year 2023**

**Editor-in-Chief and Chair of Editorial Council**  
Academician I. A. Sokolov

## IN THIS ISSUE:

SELF-TIMED PIPELINE WITH VARIABLE STAGE NUMBER <i>I. A. Sokolov, Yu. A. Stepchenkov, Yu. G. Diachenko, N. V. Morozov, and D. Yu. Diachenko</i>	4
THE PARALLEL CORPORA PERSPECTIVE ON STUDYING CONTRASTIVE PUNCTUATION <i>V. A. Nuriev and M. G. Krushkov</i>	14
INTEGRATION CAPACITIES OF SUPRACORPORA DATABASES <i>A. A. Durnovo, O. Yu. Inkova, and V. A. Nuriev</i>	24
USING A DATABASE OF STRUCTURAL TRANSFORMATIONS TO EXTRACT MULTICOMPONENT TERMINOLOGICAL UNITS <i>Yu. I. Butenko</i>	35
SEMANTIC INTERPRETATIONS OF HIGH NORMAL FORMS OF RELATIONS IN A RELATIONAL DATABASE <i>V. A. Ivanov, M. Yu. Konyshchev, S. V. Smirnov, O. V. Tarakanov, V. O. Tarakanova, and S. V. Usovik</i>	45
CLASSIFICATION PROBLEM IN CONDITIONS OF DISTORTED CAUSE-AND-EFFECT RELATIONSHIPS <i>A. A. Grusho, N. A. Grusho, M. I. Zabezhailo, A. A. Zatsarinny, E. E. Timonina, and S. Ya. Shorgin</i>	59
METHODS OF CLASSIFYING THE DISTANCE LEARNING SYSTEM USERS IN THE MODEL OF CONSTRUCTING THEIR PERSONALIZED LEARNING STRATEGIES <i>Ya. G. Martuyushova, T. A. Mineyeva, and A. V. Naumov</i>	68
EFFICIENT COMPUTATIONS IN MATRIX FACTORIZATION WITH MISSING COMPONENTS <i>M. P. Krivenko</i>	78