

ОГЛАВЛЕНИЕ

М. М. Горбунов-Посадов КАК В БИБЛИОГРАФИЧЕСКОЙ ССЫЛКЕ «НА ЛЕТУ» ФОРМИРУЕТСЯ ДАТА РАЗМЕЩЕНИЯ ПОСЛЕДНЕЙ ВЕРСИИ ПУБЛИКАЦИИ	554-565
Л. В. Городняя ОРГАНИЗАЦИЯ ВЫЧИСЛЕНИЙ И РАБОТЫ С ПАМЯТЬЮ В УЧЕБНОМ ЯЗЫКЕ ПРОГРАММИРОВАНИЯ СИНХРО	566-599
А. С. Козицын, С. А. Афонин, Д. А. Шачнев ИСПОЛЬЗОВАНИЕ ГРАФА СОАВТОРСТВА ДЛЯ ТЕМАТИЧЕСКОГО ПОИСКА КОНФЕРЕНЦИЙ ПО НАУКОМЕТРИЧЕСКИМ ДАННЫМ	600–615
О. А. Невзорова, К. С. Николаев СЕМАНТИЧЕСКОЕ АННОТИРОВАНИЕ МАТЕМАТИЧЕСКИХ ФОРМУЛ В PDF-ДОКУМЕНТАХ	616–639
Т. А. Полилова РЕЙТИНГИ БИБЛИОГРАФИЧЕСКОЙ БАЗЫ И «БЕЛЫЕ СПИСКИ»	640–670
Ю. Е. Поляк ЭТАПЫ ТРУДНОГО ПУТИ (о компьютеризации экономических исследований)	671–696
Н. И. Тарасов, В. О. Подрыга, С. В. Поляков, А. В. Тимаков ЦИФРОВАЯ ПЛАТФОРМА ДЛЯ СУПЕРКОМПЬЮТЕРНОГО МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРОЦЕССОВ НАПЫЛЕНИЯ	697–721

УДК 01

КАК В БИБЛИОГРАФИЧЕСКОЙ ССЫЛКЕ «НА ЛЕТУ» ФОРМИРУЕТСЯ ДАТА РАЗМЕЩЕНИЯ ПОСЛЕДНЕЙ ВЕРСИИ ПУБЛИКАЦИИ

М. М. Горбунов-Посадов^[0000-0002-7044-8287]

Институт прикладной математики им. М.В. Келдыша РАН, Миусская пл., 4,
Москва, 125047

gorbunov@keldysh.ru

Аннотация

Размещенную в интернете научную работу, которую ее автор постоянно поддерживает в актуальном состоянии, будем называть живой публикацией. У жанра живой публикации есть множество привлекательных черт. Однако он требует определенного расширения состава метаатрибутов публикации: наряду с традиционными атрибутами здесь на первый план выдвигается дата появления очередной, свежей редакции. Такая дата размещается на видном месте в тексте публикации. Наряду с этим весьма желательной становится включения такой динамически («на лету») формируемой даты в библиографическую ссылку на живую публикацию.

В статье рассмотрены применяющиеся сейчас способы динамического извлечения искомой даты для простой онлайн-публикации, для публикации, получившей DOI посредством Crossref, и для публикаций, размещенных в arXiv.org и preprints.org.

Ключевые слова: живая публикация, динамический компонент библиографической ссылки, свежая дата редакции, Crossref, arXiv.org, Preprints.org.

ВВЕДЕНИЕ

До последнего времени повсюду, в том числе в мире науки, господствовали пословицы «Что написано пером, не вырубишь топором» и "Littera scripta manet" («Написанная буква остается», *лат.*). Ошибку, допущенную в опубликованной статье, практически невозможно было исправить, она преследовала автора всю оставшуюся жизнь и сбивала с толку его читателей. Кроме того, получение каких-либо новых результатов в исследуемой области всякий раз требовало от автора

выпуска новой статьи, где он был вынужден весьма значительную часть текста посвятить повторению ранее опубликованных сведений, без которых свежий читатель не мог воспринять это новое содержание.

Появление интернета и последовавший за ним переход массового читателя научных публикаций в онлайн позволяют оставить эти досадные обстоятельства в прошлом. Сейчас любая статья, размещенная в онлайн, за пару минут может быть заменена ее исправленной и/или дополненной версией. Автор, не пользующийся этой открывшейся счастливой возможностью, не только лишает своего читателя шанса узнать свежие новости из заинтересовавшей его области, но и своим бездействием зачастую вынуждает его иметь дело с текстом, содержащим известную автору, но неисправленную ошибку.

Размещенную в интернете научную работу, которую ее автор постоянно поддерживает в актуальном состоянии, будем называть живой публикацией [1]. Благодаря своим очевидным преимуществам живые публикации с каждым годом приобретают все новых и новых сторонников.

••• Автор, отказавшийся от традиционной, статичной публикации в пользу живой, попадает в новую, существенно более комфортную и продуктивную среду. Допущенные ошибки и опечатки не носят теперь фатального характера, легко поддаются исправлению. Круг читателей живой публикации много шире, интерес к ней со временем нередко даже усиливается: многие читатели раз за разом возвращаются к любимившемуся тексту, не только чтобы освежить в памяти наиболее существенные моменты, но и чтобы узнать, как трансформируются взгляды автора и что нового он предложил или заметил у других в своей области.

••• Для читателя живая публикация несомненно предпочтительнее статичной. В самом деле, насколько увереннее себя чувствуешь, когда знаешь, что текст перед твоими глазами находится под неусыпным контролем автора, что в нем тщательно исправлены все неточности и ошибки, замеченные с момента первого размещения работы в онлайн, а также постоянно отслеживаются изменения, происходящие в рассматриваемой отрасли науки.

Разумеется, было бы утопией представлять жизнь автора живой публикации абсолютно безоблачной. В частности, серьезные сложности возникают здесь в

связи с обычными чиновничьими требованиями публикационной отчетности. Нередко отчетность исчисляется исключительно в штуках, и в таком случае в глазах чиновника живая публикация представляется безрассудным вызовом, подчеркнуто нерациональным расходом творческих усилий. Достаточно вспомнить, что для защиты докторской диссертации в форме научного доклада ВАК сейчас требует от 30 до 50 статей в авторитетных журналах — задача просто неподъемная в случае, если заметные усилия диссертанта отвлекаются на поддержку его ключевых, но живых и в силу этого немногочисленных публикаций.

ДАТА ПОСЛЕДНЕЙ РЕДАКЦИИ

Как читателю отличить живую публикацию от статичной? Просто добавить к ее представлению специальный значок «Публикация объявлена живой», очевидно, недостаточно. Ведь автор мог когда-то поставить этот значок и благополучно забыть и о нем, и о своем онлайн-тексте. Поэтому единственное достоверное свидетельство живой публикации — свежая дата ее последней редакции. Эта дата безусловно демонстрируется на заметном месте основного текста, например, в форме баннера, бросающегося в глаза (рис. 1).



Рис. 1. Баннер живой публикации. Указаны дата первого размещения и дата текущей (свежей) редакции

Баннер, содержащий дату, служит надежным ориентиром для читателя публикации. Такие сведения несомненно полезны. Но не менее важно оповещать читателя о том, что публикация живая, в другой часто встречающейся ситуации — при просмотре ссылки на нее в библиографическом списке.

Эта очевидная потребность в какой-то мере нашла отражение в ГОСТ Р 7.0.5-2008 [2]. Там в п. 10.4.3 введено понятие «Дата обновления» («Дата пересмотра» и т. п.), где, по-видимому, имеется в виду дата создания последней версии работы, которая была доступна автору при формировании библиографического списка [3]. Такая дата может именовать сохраненную в свое время версию, на которую в тот момент опирался автор, однако сохраненная устаревшая версия

работы вряд ли привлечет внимание читателя, которому в подавляющем большинстве случаев интересен последний, свежий материал. Предлагаемый ГОСТом библиографический список статичен, поэтому включенные в него «Даты обновления» живых публикаций со временем неизбежно теряют актуальность и перестают представлять интерес для подавляющего большинства читателей.

ДИНАМИЧЕСКИ ОБНОВЛЯЕМАЯ ДАТА

Читателю, просматривающему библиографические ссылки, разумеется, не безразлично, какие из перечисленных публикаций живые, а какие «мертвые», статические, т. е. не изменяются с момента своего первого появления. В то же время, и здесь недостаточно просто пометить публикацию в библиографическом списке как «живую»: автор мог много лет назад забыть о своем намерении обратиться к данному жанру, и тогда такая пометка потеряет смысл, будет просто дезориентировать читателя. И здесь единственное надежное свидетельство «живости» публикации — предъявление онлайн-читателю свежей даты ее последней редакции.

Для реализации такого предъявления желателен формат HTML. Если относительно недавно основным и практически единственным средством онлайн-представления научной публикации был устаревающий формат PDF, то сейчас HTML благодаря своим многочисленным преимуществам постепенно завоевывает в этой сфере устойчивые позиции. Одно из преимуществ HTML — возможность относительно несложно включать в текст онлайн-публикации динамически формируемые элементы.

Так, чтобы привлечь внимание читателя библиографического списка к живым публикациям, для HTML были разработаны программные инструменты [4, 5], реализующие кроссдоменные связи и позволяющие с их помощью динамически дополнить привычный текст библиографической записи новым важным компонентом — свежей датой последней редакции публикации. Для этого при формировании текста библиографической записи, предъявляемого онлайн-читателю, выполняется обращение к живой публикации, размещенной, вообще говоря, в другом домене. Дата обновления хранится там в определенном формате в атрибуте живой публикации, извлекается оттуда «на лету» и включается в онлайн-текст.

Тем самым онлайн-читатель библиографического списка всегда видит действительно последнюю на текущий момент дату появления очередной версии живой публикации. Для оформления в библиографической ссылке динамической даты появления будем использовать префикс «Доступна редакция от», а саму дату окружать символами «≈», например,

Горбунов-Посадов М.М. Живая публикация // Открытые системы. 2011, № 4. С. 48–49.
Доступна редакция от ≈ 2022-07-03 ≈
<https://keldysh.ru/gorbunov/live.htm>

Гиперссылка, включаемая здесь в библиографическую запись, ведет именно на свежую редакцию публикации.

CROSSREF

Гиперссылка, включаемая в библиографическую запись, из-за хорошо известных технологических соображений сейчас все чаще ведет не непосредственно к файлу публикации, а отсылает к DOI (Digital Object Identifier), который в большинстве случаев поставляется агентством Crossref. Crossref долгое время занимало жесткую позицию по отношению к живым публикациям: в правилах агентства было записано, что в публикацию, получившую DOI, впоследствии нельзя вносить никаких изменений.

Однако со временем живые публикации в Crossref были в полной мере легализованы. С одной стороны, из правил было исключено требование неизменности материала, получившего DOI. С другой стороны, был реализован механизм Crossmark, предполагающий, что все вновь появляющиеся версии живой публикации сосуществуют со своими предшественницами, и каждая из них получает собственный DOI. В каждой из этих версий на видном месте помещается обслуживающий живую публикацию значок "Check for updates" (рис. 2).

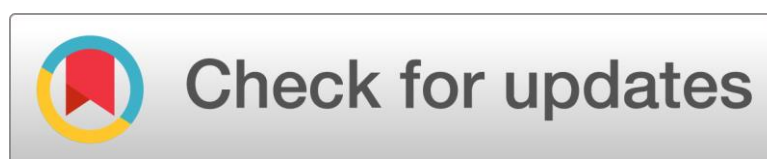


Рис. 2. Значок Crossmark, обслуживающий живую публикацию в Crossref.

Crossmark не позволяет включить непосредственно в библиографическую ссылку интересующую нас постоянно действующую гиперссылку на свежую версию живой публикации. Вместо этого в библиографическую ссылку здесь предлагается включать DOI-гиперссылку на текущую, устаревающую со временем версию. Однако читатель, как-либо забредший на возможно устаревшую версию, кликнув по Crossmark, может узнать, существует ли и где расположена более свежая версия материала, а также не ретрагирован (т. е. не отозван) ли этот материал редакцией.

Механизм обслуживания живой публикации Crossmark, предлагаемый Crossref, представляется нерациональным. Доступ к свежей версии через устаревшую противоестественен. Кроме того, читатель может просто не обратить внимания на значок Crossmark и тем самым не догадаться о существовании более свежей версии публикации. Наконец, даже если бы как-то удалось узнать дату последней редакции свежей версии, включить эту дату в библиографическую ссылку на устаревшую версию нельзя — читатель может ошибочно принять ее за дату редакции устаревшей версии. Однако без указания свежей даты ссылка на живую публикацию скучна, а в определенном смысле даже некорректна.

В то же время Crossref не настаивает на единственности подхода Crossmark к обслуживанию версий: допускается не только свой DOI для каждой версии (как в Crossmark), но и один DOI для всех версий, которые в таком случае сменяют друг друга под этим адресом. Каждый из подходов, по мнению Crossref [6], имеет свои преимущества и недостатки.

На наш взгляд, предпочтение следует отдать подходу с единственным DOI. Продуктивнее и комфортнее для читателя поставить дело так, чтобы внешняя ссылка, в данном случае DOI-гиперссылка, всегда вела непосредственно на позднейшую, самую свежую версию материала. Иначе говоря, пусть очередные версии живой публикации сменяют друг друга под одним и тем же DOI, а для любителя старины на странице живой публикации можно предусмотреть отсылку к формируемому и сохраняемому где-то в стороне протоколу происходивших изменений живой публикации.

При такой организации обслуживания соответствующую библиографическую ссылку можно и нужно дополнить динамически обновляемой датой последней редакции живой публикации [7] — надежным свидетельством того, что автор не забывает о постоянной поддержке своей работы. Например,

Горбунов-Посадов М.М. Библиографическая ссылка на онлайн-публикацию // Препринты ИПМ им. М.В. Келдыша. 2020. № 11. 28 с. Доступна редакция от ≈ 2022-07-19 ≈
<https://doi.org/10.20948/prepr-2020-11>

Если автор планирует каждую новую версию живой публикации заново индексировать (разумеется, под тем же DOI) в Crossref, то он может не объявлять дату в файле живой публикации: если дата в файле не объявлена, то в библиографическую запись будет вставлена дата последнего индексирования публикации в Crossref.

В то же время, если внесенные изменения коснулись только основного текста, т. е. ни место размещения файла, ни библиография, ни аннотация, ни другие метаатрибуты не изменились, то можно, вообще говоря, немного сэкономить усилия — очередную версию разместить, но не индексировать в Crossref заново. В этом случае для обслуживания динамической даты, включаемой в библиографическую ссылку, в файле публикации потребуется явно указать дату последней редакции, как и в случае прямой (без DOI) ссылки на живую публикацию.

ARXIV

Как известно, arXiv [8] — крупнейший архив научных статей, размещаемых в открытом доступе. arXiv создан и поддерживается Корнельским университетом (США). Функционирует с 1991 года, к 2022 году в arXiv было размещено свыше двух миллионов статей.

arXiv поддерживает размещение живых публикаций. Автор вправе в произвольные моменты времени размещать в arXiv новые и новые версии своей живой статьи, файлы которых получают адреса (URL) с суффиксами v1, v2, v3, ... (рис. 3).

Submission history

From: LVC Publications [[view email](#)] [via Lvc Publications as proxy]

[v1] Thu, 5 Oct 2017 19:18:51 UTC (4,609 KB)

[v2] Wed, 11 Oct 2017 16:26:26 UTC (4,608 KB)

[v3] Tue, 7 Nov 2017 22:45:22 UTC (4,608 KB)

[v4] Tue, 8 Oct 2019 13:29:33 UTC (4,507 KB)

Рис. 3. Фрагмент карточки <https://arxiv.org/abs/1710.02185>:
четыре версии живой публикации.

Для доступа к последней (свежей) версии живой публикации в arXiv используется URL без суффикса, например, <https://arxiv.org/abs/2103.10761>. Библиографическую ссылку, использующую такой укороченный URL, можно дополнить динамически обновляемой датой размещения в arXiv последней версии [9], например,

Gorbunov-Posadov M.M. Alive publication.

Revision from ≈ 2021-03-19 ≈

<https://arxiv.org/abs/2103.10761>

В таком случае объявлять дату последней редакции в файле публикации не требуется: дата извлекается из системных данных arXiv.

В 2022 году статьи в arXiv получили DOI. DOI здесь один для всех версий статьи и ведет к самой свежей, последней из имеющихся версий. По-видимому, теперь ссылаться на статью имеет смысл именно посредством DOI, например, <https://doi.org/10.48550/arXiv.2103.10761>. Разумеется, и тут ссылку желательно дополнить автоматически обновляемой датой размещения последней версии.

Отметим, что все фигурировавшие выше адреса версий ведут не на полный текст, а на карточку статьи в arXiv (рис. 3), что доставляет читателю некоторые неудобства — до полного текста здесь удастся добраться только не менее чем за два клика. Однако таким образом в arXiv удастся добиться того, что даже читатель, пришедший по URL конкретной, возможно устаревшей версии, скорее всего заметит на карточке несколько дат размещения версий и сообразит, что

статья живая, т. е. наряду с файлом версии, указанной в исходном URL, ему сейчас доступна и последняя, самая свежая версия.

PREPRINTS

Preprints [10] — относительно небольшой, но популярный среди российских авторов архив научных статей, размещаемых в открытом доступе. Preprints поддерживается издательством MDPI, специализирующимся на издании журналов открытого доступа. Preprints функционирует с 2017 года, к 2022 году в архиве было размещено более 30 тысяч статей. Архив междисциплинарный, размещает статьи из всех областей научных исследований.

Как и arXiv, Preprints поддерживает живые публикации. Как и в arXiv, статьи получают суффиксы v1, v2, v3, ..., а для доступа к последней версии живой публикации используется URL без суффикса, например, <https://preprints.org/manuscript/202209.0202>.

В то же время отношения Preprints с DOI абсолютно несхожи с arXiv. Здесь каждая версия статьи получает свой DOI. DOI последовательных версий статьи получают те же суффиксы v1, v2, v3, ... И, главное, Preprints почему-то не поддерживают DOI без суффикса для доступа к последней версии статьи. Тем самым автор, желающий включить в библиографическую ссылку на Preprints автоматически обновляемую дату последней версии статьи, вынужден работать не с DOI, а с коротким URL:

Gorbunov-Posadov M.M. Dynamically updated alive publication date // Preprints, 2022. Revision from ≈ 2022-11-07 ≈.
<https://preprints.org/manuscript/202209.0202>

ЗАКЛЮЧЕНИЕ

Живая публикация в силу своих очевидных преимуществ неуклонно завоевывает все новые и новые позиции в научном сообществе. Когда чиновник, принимающий организационные решения, наконец признает весомой научной заслугой автора не только публикацию статьи в авторитетном издании, но и поддержку в актуальном состоянии получившей признание живой публикации, тогда

распространение живых публикаций приобретет массовый характер, что несомненно пойдет на пользу современной науке.

Включение в библиографическую ссылку динамической даты последней редакции живой публикации требует от автора относительно небольших усилий. Для читателя же такая дата оказывается чрезвычайно полезной, переходы из библиографического списка к недавно обновленным живым публикациям выполняются обычно в разы чаще, чем к статичным, не снабженным указанием динамической даты.

СПИСОК ЛИТЕРАТУРЫ

1. Горбунов-Посадов М.М. Живая публикация // Открытые системы. 2011. № 4. С. 48–49. <https://keldysh.ru/gorbunov/live.htm>
2. ГОСТ Р 7.0.5-2008 БИБЛИОГРАФИЧЕСКАЯ ССЫЛКА. Общие требования и правила составления. <https://docs.cntd.ru/document/1200063713>
3. Горбунов-Посадов М.М. Библиографическая ссылка на онлайн-публикацию // Препринты ИПМ им. М.В. Келдыша. 2020. № 11. 32 с. <https://doi.org/10.20948/prepr-2020-11>
4. Горбунов-Посадов М.М., Скорнякова Р.Ю. Обновляемая дата последней редакции в ссылке на живую публикацию // Препринты ИПМ им. М.В. Келдыша. 2017. № 82. 14 с. <https://doi.org/10.20948/prepr-2017-82>
5. Живая публикация. Инструменты поддержки. <https://alive.keldysh.ru/>
6. Crossref. Preprint metadata recommendations. <https://www.crossref.org/pdfs/preprint-ag-recommendations-2022-07.pdf>
7. Crossref: DOI-ссылка на живую публикацию. <https://alive.keldysh.ru/crossref.htm>
8. arXiv: open-access e-print archive. <https://arxiv.org/>
9. arXiv: живая публикация в arxiv.org. <https://alive.keldysh.ru/arxiv.htm>
10. Preprints: the multidisciplinary preprint platform. <https://preprints.org>

HOW THE LATEST RELEASE DATE OF PUBLICATION IS FORMED IN BIBLIOGRAPHIC REFERENCE "ON THE FLY"

M.M. Gorbunov-Posadov^[0000-0002-7044-8287]

Keldysh Institute of Applied Mathematics, Miusskaya sq., 4, Moscow, 125047, Russia
gorbunov@keldysh.ru

Abstract

The scientific work posted on the Internet, which its author constantly keeps up to date, will be called an alive publication. The genre of alive publishing has many attractive features. However, it requires a certain expansion of the composition of the meta-attributes of the publication: along with the traditional attributes, the date of the appearance of the new, fresh edition is brought to the fore here. Such date is placed in a prominent place in the text of the publication. Along with this, it becomes highly desirable to include such a dynamically ("on the fly") generated date in a bibliographic link to an alive publication. The currently used methods of dynamic extraction of this date are considered for a simple online publication, for a publication that has received a DOI through Crossref, and for a publication posted in arXiv.org.

Keywords: *alive publication, dynamic component of the bibliographic reference, the latest revision date, Crossref, arXiv.org*

REFERENCES

1. Gorbunov-Posadov M.M. Zhivaia publikatsiia // Otkrytye sistemy. 2011. № 4. S. 48–49. <https://keldysh.ru/gorbunov/live.htm>
2. GOST R 7.0.5-2008 BIBLIOGRAFICHESKAIa SSYLKA. Obshchie trebovaniia i pravila sostavleniia. <https://docs.cntd.ru/document/1200063713>
3. Gorbunov-Posadov M.M. Bibliograficheskaiia ssylka na onlainovuiu publikatsiiu // Preprinty IPM im. M.V. Keldysha. 2020. № 11. 32 s. <https://doi.org/10.20948/prepr-2020-11>
4. Gorbunov-Posadov M.M., Skorniakova R.Iu. Obnovliaemaia data poslednei redaktsii v ssylke na zhivuiu publikatsiiu // Preprinty IPM im. M.V. Keldysha. 2017. № 82. 14 s. <https://doi.org/10.20948/prepr-2017-82>
5. Zhivaia publikatsiia. Instrumenty podderzhki. <https://alive.keldysh.ru/>

6. Crossref. Preprint metadata recommendations.

<https://www.crossref.org/pdfs/preprint-ag-recommendations-2022-07.pdf>

7. Crossref: DOI-ssylka na zhivuiu publikatsiiu.

<https://alive.keldysh.ru/crossref.htm>

8. arXiv: open-access e-print archive. <https://arxiv.org/>

9. arXiv: zhivaia publikatsiia v arxiv.org. <https://alive.keldysh.ru/arxiv.htm>

10. Preprints: the multidisciplinary preprint platform. <https://preprints.org>

СВЕДЕНИЯ ОБ АВТОРЕ



ГОРБУНОВ-ПОСАДОВ Михаил Михайлович – главный научный сотрудник ИПМ им. М.В. Келдыша РАН, д. ф.-м. н.

Mikhail Mikhailovich GORBUNOV-POSADOV – Keldysh Institute of Applied Mathematics, chief researcher. Current scientific interests: programs extensibility, scientific publications.

email: gorbunov@keldysh.ru;

ORCID: 0000-0002-7044-8287

Материал поступил в редакцию 23 декабря 2022 года

УДК 004.43 (042.4)

ОРГАНИЗАЦИЯ ВЫЧИСЛЕНИЙ И РАБОТЫ С ПАМЯТЬЮ В УЧЕБНОМ ЯЗЫКЕ ПРОГРАММИРОВАНИЯ СИНХРО

Л. В. Городня^{1,2} [0000-0002-4639-9032]

¹Институт систем информатики им. А.П. Ершова СО РАН, 630090

г. Новосибирск, проспект Академика Лаврентьева, 6

²Новосибирский государственный университет, 630090, Новосибирск,
ул. Пирогова, 1

¹gorod@iis.nsk.su

Аннотация

Статья посвящена ряду решений, принятых в проекте разрабатываемого в Лаборатории информационных систем Института систем информатики СО РАН учебного языка программирования СИНХРО, предназначенного для ознакомления с базовыми явлениями взаимодействия процессов и управления вычислениями над общей памятью. В центре внимания находится парадигма функционального программирования. Язык ориентирован на школьников младших и средних классов, а также студентов младших курсов и непрофессионалов. При обучении используется опыт оперирования игрушечными роботами, перемещающимися на клетчатой доске. Статья представляет интерес для всех, кто интересуется проблемами современной информатики, программирования и информационных технологий, особенно проблемами параллельных вычислений на суперкомпьютерах и распределённых системах, и вообще применением многопроцессорных комплексов.

Ключевые слова: учебные языки программирования, виртуальная машина, система команд, функциональное программирование, восстановление данных, освобождение памяти, многопоточные программы, параллельные вычисления, общая память, взаимодействие процессов.

ВВЕДЕНИЕ

Системе подготовки программистов предстоит освоить мир параллелизма и создать методику опережающего ознакомления с его не вполне очевидными явлениями, что и побудило к эксперименту по разработке учебного языка программирования СИНХРО [1]. Алгоритмы становятся параллельными, программы – многопоточными, компьютеры – многопроцессорными. Очередное действие может начать выполняться до завершения предыдущего, средства управления вычислениями кроме логических значений используют события и многое другое [2–6], что приводит к размежеванию синтаксиса и семантики параллельных вычислений на уровне выражений [7–9]. И всё это происходит в многоязыковой обстановке – мощность пространства языков программирования давно перевалила за десятки тысяч [10, 11].

Первая цель опережающего ознакомления с параллелизмом – профилактика жесткого привыкания к принципам самого распространённого однопроцессорного, последовательного, императивно-процедурного программирования, осложняющего овладение многопоточным программированием [12, 13]. Следует признать, что программирование и его техническая основа в последние десятилетия претерпели значительные изменения. Вторая цель – приобретение опыта применения механизмов, актуальность которых растёт в связи с усложнением ИТ и расширением классов сложных задач, обременённых трудно удостоверяемыми требованиями, такими как живучесть, надёжность и безопасность программных инструментов.

Изложение начинается с нестроого описания основных механизмов языка СИНХРО (раздел 1) и обоснования решений по созданию и описанию учебного языка программирования языка СИНХРО (раздел 2). Затем описаны особенности семантики языка (раздел 3) и прагматики перехода от потоков к процессам (раздел 4), включая вытекающие из этого решения по системе команд виртуальной машины, допускающей поддержку взаимодействия процессов (раздел 5), и описание механизма взаимодействия общей и локальной памяти (раздел 6). В заключении даны выводы по образовательным проблемам параллельного программирования.

1. ОБЩИЕ ПОЛОЖЕНИЯ

Детское программирование давно уже не редкость [14–17], и при оценке образовательного значения парадигм программирования признают фундаментальными – функциональное, параллельное и императивно-процедурное, а логическое и ООП рассматривают как дополнительные парадигмы. Опережающее освоение параллелизма пока не получило признания, хотя реализация такого подхода возможна на базе учебных и новых языков программирования благодаря присущей им мультипарадигмальности.

Как это типично для учебных языков программирования, язык СИНХРО поддерживает как основные, так и фундаментальные парадигмы программирования, включая параллельное программирование и метапрограммирование. Использование разных парадигм приводит к использованию разных форм представления многопоточных программ, что методически обусловлено профилактикой привыкания к одной форме, которая может тормозить переход к новым формам (Эффект второй системы) [18]. Эти формы семантически эквивалентны, они порождают одно и то же семейство процессов.

Проект учебного языка СИНХРО рассчитан на формирование интуитивных моделей в процессе ознакомления с основными явлениями параллельного программирования, встречающимися в учебно-методических и научных материалах, языках высокого уровня (ЯВУ), языках высокопроизводительного программирования и механизмах сетевых информационных сервисов, образующих инструменты современной ИТ-индустрии. Такие модели необходимы для успешного обучения в любой практической области, в частности, для перехода к производственной деятельности и обеспечению надёжности и безопасности ИТ.

Следует обратить внимание, что при обучении не всегда более простую сущность изучить легче, чем более сложную. Реально как простое воспринимается знакомое и привычное, а новое и непривычное кажется более сложным, даже если оно математически и конструктивно много проще.

Основные понятия языка СИНХРО – фрагменты выражений, действий, вычислений и данных, включая процессоры, программы, языки программирования или системы. В учебном процессе язык предстаёт как череда диалектов, соответствующих фазам обучения.

По умолчанию обстановка для выполнения программы независимо от диалекта содержит систему программирования на полном входном языке. Определение программы содержит контекст, в котором синхронизация набора потоков корректна, иначе программа выполняется до обнаружения невыполнимого действия и приостанавливается.

2. ВЫБОР РЕШЕНИЙ

Определение языка СИНХРО потребовало заметного числа неочевидных решений.

2.1. Учёт скрытой грамматики деятельности. Обучение параллельному программированию требует пересмотра методик обучения для формирования скрытой грамматики деятельности по организации процессов [19].

2.2. Выделение диалектов по целям обучения. Мульти – ознакомление с феноменами параллелизма; Трансформ – опыт метапрограммирования; Асинхр – приобретение навыков подготовки многопоточных программ; Синпар – изучение методов представления взаимодействующих процессов; АМК – эксперименты по многопроцессорному программированию. Диалекты можно использовать автономно и совместно.

2.3. Использование мультипарадигмальности. Стихийно сформированные интуитивные модели организации процессов недостаточно удобны при обучении функциональному, параллельному и логическому программированию потому, что в обычной жизни прецеденты применения таких парадигм встречаются не часто и обычно рассредоточено по времени – нет компактности [19].

2.4. Приоритет параллелизму. Предварительное освоение иллюстративного материала для целенаправленного формирования интуитивных моделей, поддерживающих парадигму деятельности по организации параллельных процессов [1, 5, 7, 12, 20].

2.5. Ведущая парадигма функционального программирования, освобождающая для начала от проблем с побочными эффектами и ошибок, сложно зависящих от реального времени.

2.6. Разнообразие архитектур. Многие идеи организации процессов на уровне операционных систем слишком ограничены механизмами параллельной обработке на базе очередей и векторов [21].

2.7. Методическая обусловленность введения понятий программирования. Учтены идеи известных учебных языков программирования (Basic, Pascal, Logo, Grow, Karel, Робик, A++, Oz). Акцент на опыт методики введения понятий программирования на базе языка начального обучения программированию Робик [13].

2.8. Взаимодействующие процессы. Энтони Хоар (Antony Hoare) давно отметил, что параллелизм – вызов интеллекту человека. В его книге «Взаимодействующие последовательные процессы» [5] показано, что параллельные композиции внешне не сложнее последовательных. Именно таким образом простота пользовательских интерфейсов маскирует истинную сложность программ и процессов вычислений.

2.9. Расширяемость языка и системы программирования. Многие понятия языка программирования при переходе к параллельному программированию претерпевают изменения, и возникает необходимость в дополнительных понятиях, расширяющих ядро языка на уровне оболочки или отдельной программы. При определении языка СИНХРО использованы идеи языков Lisp, F#, C#, допускающих динамическую обработку кода программ, дополненные гибридными решениями по обработке памяти.

2.10. Ленивые вычисления и мемоизация. Выполнение параллельных процессов часто требует независимости порядка вычислений от последовательности представления действий в программе. Отчасти это делается с помощью замыкания функций и мемоизации, сохраняющей соответствие между значениями параметров и полученных результатов.

2.11. Трансформационная семантика. Общие решения по обеспечению лаконизма, универсальности, конструктивности и расширяемости приводят к трансформационной семантике языка программирования. Это является естественным полигоном для метапрограммирования как основы для оптимизации программ, признаваемой как основная проблема разработки компиляторов [22].

2.12. Просачивание операций и функций. Результат операции над скалярами в языках параллельного программирования обычно автоматически распространяется на произвольные однородные структуры данных. Этот механизм в языке СИНХРО распространяется на технику применения функций, что повышает лаконизм выражений [1, 23].

2.13. Фильтрация данных. Результат фильтрации исчезает из аргумента – он переносится в другую структуру данных [23–25], что удобно при подготовке программ при решении задач методом перебора.

2.14. Фрагменты программ и синтаксическое подобие. Для задания синтаксического подобия фрагментной переменной её значениям используется понятие вида переменной. При определении макросов бинарная операция « $\sim\sim$ » задаёт вид левого аргумента с помощью строки, являющейся правым аргументом. Все вхождения фрагментной переменной в схему программы должны быть синтаксически эквивалентны её вхождению в строку, задающую вид фрагментной переменной [26, 27].

2.15. Метапрограммирование для преобразования программ. В случае многопоточных программ возможно преобразование потоков, нацеленное на сведение к однородной системе. Обратимость преобразований и чувствительность их результата к информационным связям между фрагментами программы даёт умение формализовать критерии применимости трансформаций и выбора подходящего варианта [22].

2.16. Взаимодействие локальной и общей памяти. Предложена гибридная реализация императивной синхронизации взаимодействия локальной и общей памяти, совмещающая возможности «сборки мусора» и счетчика указателей на элементы памяти. [21, 28, 29].

2.17. Клетчатое поле. Дано предпочтение задачам, допускающим иллюстрацию решений в терминах перемещения по клетчатой доске [13].

2.18. Относительное описание языка программирования. Описание языка представляет собой череду диалектов, каждый из которых определён как дополнение к предыдущему диалекту.

3 МНОГОПОТОЧНЫЕ ПРОГРАММЫ

«Не верь глазам своим!»

Конкретный свод решений по семантике языка СИНХРО связан с классом рассматриваемых многопоточных программ.

3.1. Локализация имен используется лишь при определении исполнителей, потоков и функций. Нет иерархии и статической типизации.

3.2. Многопоточная программа представляется как набор потоков действий и может быть размечена барьерами на слои. Можно задавать внутреннюю и внешнюю очереди барьеров, определяющие порядок выполнения помеченных барьерами слоёв и действий [5].

3.3. Достижение барьера рассматривается как событие. Событие не сбрасывается, пока все ждущие его потоки не будут готовы или их выполнение не будет оценено как невозможное. [6].

3.3. Выражения могут использовать размещённые в общей памяти данные, но не изменяют их. Директивы могут изменять хранимые в общей памяти данные [1, 28, 29].

3.4. Структуры данных в программе наполняются элементами, порядок вычисления которых может отличаться от порядка вхождения в программу [28, 29]. Роль такого синтаксического отделения показана на примерах в таблице 1. Различие проявляется в мощности семейства допустимых процессов вычисления, что можно видеть на примерах (табл. 2), демонстрирующих такую разницу на простых выражениях со списками и векторами, воспринимаемыми зрительно как равнозначные, содержащие одинаковые элементы, но семантически они различны.

В первом примере согласно порядку записи в программе вычисляется a' , затем b' . После этого в памяти размещается $\downarrow b$, получается список (b), затем в этом списке размещается $\downarrow a$ и получается список (a b). Во втором примере b' может быть вычислено раньше, чем a' , и его результат может быть сразу или позже размещён $\downarrow b$ в списке (b) до вычисления a' и размещения полученного результата $\downarrow a$. Это даст такой же список (a b). Выражения «(a, b)» и «(a; b)» функционально эквивалентны, а мощность пространств допустимых процессов вычислений для них отличается.

Таблица 1. Порядок вычислений не зависит от порядка доступа к данным

№	Выражение	Пояснение
1	'('Выр (';' Выр)...)'	Список последовательно вычисляемых элементов, заполняемый в порядке записи в программе.
2	'('Выр (',' Выр)...)'	Список, заполняемый элементами в порядке записи в программе, вычисляемыми асинхронно, возможно в другой последовательности.
3	'['Выр (',' Выр)...)'	Вектор, заполняемый элементами в порядке записи в программе, вычисляемыми асинхронно, возможно в другой последовательности.
4	'['Выр('; ' Выр)...)'	Вектор, заполняемый последовательно вычисляемыми элементами в порядке записи в программе.

В третьем примере размещение $1:\downarrow a$ первым элементом вектора может произойти как раньше вычисления b' , так и позже, результат которого размещается вторым элементом вектора $2:\downarrow b$, а размещение $1:\downarrow a$ может быть выполнено после размещения $2:\downarrow b$. В четвертом примере допускается возможность вычисления b' раньше вычисления a' . Результат совпадает с результатом третьего примера, различна лишь мощность пространств допустимых процессов вычисления. Выражения « $[a; b]$ » и « $[a, b]$ » функционально эквивалентны. При ясном понимании теоретической разницы, на лабораторной практике многие студенты учитывают допустимые процессы лишь первого и третьего примеров, глаза воспринимают текст как последовательность независимо от разделителя. В задачу разработки языка СИНХРО входит поддержка лабораторных работ для формирования навыков учёта полных семейств допустимых процессов. Многопоточная программа допускает асинхронное выполнение потоков и действий, что означает в соответствии с принципом равноправия параметров функций независимость порядка вычисления выражений от порядка их вхождения в структуры данных, рассматриваемые как функции над этими выражениями.

Таблица 2. Разница в числе допустимых процессов вычисления выражений с разными структурами данных

№	Выражение	Процессы вычисления	Пояснение
1	(a; b)	{ a' b' ↓b ↓a }	Один допустимый процесс
2	(a, b)	{ a' b' ↓b ↓a b' a' ↓b ↓a b' ↓b a' ↓a }	Три варианта допустимых процессов
3	[a; b]	{ a' b' 2:↓b 1:↓a a' 1:↓a b' 2:↓b a' b' 1:↓a 2:↓b }	Три варианта допустимых процессов
4	[a, b]	{ a' b' 2:↓b 1:↓a a' 1:↓a b' 2:↓b a' b' 1:↓a 2:↓b b' 2:↓b a' 1:↓a b' a' 1:↓a 2:↓b b' a' 2:↓b 1:↓a }	Шесть вариантов допустимых процессов

где x' – вычисление выражения x , $\downarrow x$ – размещение полученного значения x в памяти, $n:\downarrow x$ – размещение значения x как элемента вектора с номером n , { ... } – множество допустимых процессов, | – разделитель элементов множества процессов.

3.5. Навыки предвидеть проблемы. В задачи ознакомления с параллелизмом входит научиться обнаруживать и предвидеть опасности, перестраивать и отлаживать программы при обнаружении неудачных взаимодействий потоков. Директива при благополучном исходе обработки памяти даёт результат подобно выражению.

3.6. Условия выполнения действий. При отладке сценариев можно задавать ориентировочное время срабатывания отдельных действий. Условное выполнение команды приводит к сбросу соответствующего сигнала. Действия, связанные с изменением состояния памяти, подчинены механизму транзакций, т. е. признание их безуспешными влечёт восстановление памяти в состояние, предшествующее этому действию.

3.7. Определение исполнителя выглядит как сценарий его функционирования, запускаемый при вызове исполнителя. При исполнении сценария ведется

учет частоты выполнения вероятностных действий как в системах для разработки компьютерных игр.

3.8. Постановка учебной задачи для экспериментов с параллельными алгоритмами на языке СИНХРО формулируется в терминах оперирования исполнителями [13].

3.9. Действия допускают схемы обхода, выглядящие подобно оператору IF без ELSE.

3.10. Потоки могут отлаживаться независимо друг от друга в предположении, что каждый из них можно располагать в отдельном контексте, их могут выполнять разные процессоры. Кроме того, приняты следующие ограничения:

- взаимодействия потоков выполняются только через синхронизацию;
- одновременно исполняемые потоки могут обмениваться данными через общую память;
- поддерживается восстановление многократно выполняемых фрагментов.

4. ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ: ОТ ПРЕДСТАВЛЕНИЯ ПОТОКОВ К ОРГАНИЗАЦИИ ПРОЦЕССОВ

При ознакомлении с параллелизмом полезно несколько расширить свод базовых понятий программирования, чтобы допускать разнообразие возможных учебных задач, позволяющих сформировать интуитивную картину многопоточности. В однопроцессорном программировании нет необходимости в представлении процессора – он существует по умолчанию всегда. Во многих учебных задачах для ознакомления с параллелизмом в роли процессоров выступают разные роботы, учебно-игровые исполнители, программируемые устройства, многопроцессорные конфигурации, поэтому представление процессоров или выполняемых ими процессов становятся столь же необходимым данным как представления значений и функций. Можно вспомнить, что в UNIX унифицировано представление файлов, устройств, заданий и процессов.

В языке РОБИК использовался образ фабрики роботов, которых можно было придумывать, тиражировать, именовать, выключать или включать в обстановку по мере развития сюжета учебной игры или оперирования программой. В языке trC используются номера процессоров. Сети управления взаимосвязанными по-

токами – не более чем усложнённая структура данных, допускающая равноправное расположение данных, действий и процессоров в узлах сети, примерно, как это проектировалось в языке БАРС. Система управления выполнением действий учитывает геометрию сети, допускающую синхронизацию и преобразования фрагментов программ при их сборке, отладке и оптимизации, что образует полигон для учебных задач по мета программированию. На так расширенном своде понятий можно предлагать новые учебные задачи, включая оттеснение маловероятных вычислений, балансировку нагрузки процессоров, представление пространства итераций. Возможны эксперименты по автоматизированному и ручному распараллеливанию программ, моделированию необходимой при отладке идентичности повторных прогонов на суперкомпьютерах и представлению много контактных узлов для достижения эффективного использования нестандартных аппаратных решений.

Пришло время признать, что при переходе к производственным программам и параллельным вычислениям удобство приложения, работоспособность и производительность программ на практике оцениваются выше, чем полнота отладки и повышение эффективности, мало заметные пользователю. Корректность параллельных программ часто наследуется от ранее созданных последовательных программ решения тех же самых задач. Параллелизм привлекается лишь ради ускорения вычислений. Массово применяется изрядное число жизнеспособных или работоспособных программных систем, корректность которых сомнительна, что нисколько не мешает их популярности. Известны и примеры резкого взлёта удобных языков программирования, существенно проигрывающих в эффективности привычному семейству C/C++/C#.

На практике работает иное понимание правильности, обусловленное ожиданиями или привычками пользователя. Оно не формализовано и не вполне согласуется с общепринятыми и научно обоснованными рекомендациями, нацеливающими на эффективность или производительность, а также на формальную верификацию, хотя верификация заметно прогрессирует в рамках парадигмы функционального программирования и со временем может смягчить груз ответственности программиста за качество программных продуктов. Возникает специ-

альная задача обеспечивать успех и производительность программ без потери достоинств формальной верификации и системных решений, нередко дающих улучшение, значительно превосходящее теоретические прогнозы.

Обычно в системы производственного функционального программирования включают механизмы практического компромисса, внешне выглядящие как специальные функции, не нарушающие стиль записи программ. Например, Lisp 1.5 [30], Clisp, Stmcl, Clojure и другие члены семейства Lisp предоставляют варианты функций, обладающих разными свойствами при сохранении функциональной эквивалентности. В их числе контроль типов данных, схемы циклов, возможность восстановления данных, программируемые прогнозы времени счёта и объёма расходуемой памяти, мемоизация, обеспечивающая многократное использование результатов вычислений, и псевдофункции, дающие доступ к разным устройствам, начиная со средств ввода-вывода и обработки файлов.

По умолчанию виртуальная машина языка СИНХРО имеет хотя бы один процессор, на котором происходит выполнение основной программы. Компилятор по многопоточной программе, представленной средствами высокого уровня, строит функционально эквивалентную ей многопроцессорную программу низкого уровня, выполняемую на виртуальной машине. Комплект потоков при компиляции преобразуется в комплекс процессов, каждый из которых обычно выполняется одним процессором, если не возникает проблем неравномерной загрузки процессоров. Система команд виртуальной машины выбрана из соображений удобства конструирования шаблонов компиляции функций и операций языка СИНХРО. Предполагается, что ознакомление с параллелизмом включает упражнения на конструирование шаблонов, ориентированных на разные многопроцессорные конфигурации.

Контекст выполнения программы при переходе к процессам становится общей памятью, данные из неё доступны всем процессам многопроцессорной программы. Общая память содержит представления имён и значений глобальных переменных, возможно изменяемых разными процессами с помощью так называемых «деструктивных» функций, работающих подобно привычным присваиваниям. Различие в том, что каждая деструктивная функция обладает чистым функциональным эквивалентом, позволяющим отладку выполнять в два шага. Сначала

безопасная отладка под защитой принципа неизменяемости данных, затем – повышение эффективности выборочной заменой чистых функций на их деструктивные аналоги. Память состоит из двух частей – блока регистров прямого доступа к глобальным переменным и кучи произвольных структур данных для хранения значений переменных с протоколами, содержащими старые значения после присваивания на случай необходимости восстановления состояний памяти.

Для простоты изложения здесь не рассматривается учет разнообразия категорий систем команд отдельных процессоров и видов используемой памяти с различной дисциплиной функционирования, хотя в учебных задачах всё это имеет место. Механизмы освобождения памяти обычно требуют приостановки вычислений, что снижает общую производительность программы. Это особенно заметно при необходимости освободить общую память, приостанавливая все процессы комплекса. Возможные подходы к смягчению этой проблемы рассмотрены в разделе 6.

5. ВИРТУАЛЬНАЯ МАШИНА АМК

При функционировании виртуальной машины, в отличие от реальной, число процессоров может изменяться, что влечёт появление динамических запросов к памяти, не всегда заметных при статическом анализе и компиляции программы. Кроме обычных процессоров в выполнении программы могут участвовать дополнительные устройства, которые можно рассматривать как специальные процессоры, способные выполнять разные действия по запросам обычного процессора. Каждый обычный процессор выполняет один процесс, порождаемый одним рядом команд. Шаги разных процессов не синхронизованы, но могут происходить одновременно.

Учитывая особенности компиляции языков, поддерживающих парадигму функционального программирования, при описании системы команд ядра языка СИНХРО используем расширение машины SECD, предложенной П. Ландиным для спецификации аппаратной реализации языка Lisp. Машина работает над регистрами, приспособленными к хранению списков. При определении виртуальной машины языка СИНХРО к регистрам системы команд «SECD» добавлен регистр «m» (Memory) для описания команд над общей памятью. Состояние процесса

полностью определяется содержимым этих регистров. Система команд виртуальной машины поддерживает выполнение результата компиляции функций и операций языка СИНХРО. Часть команд определена в книге П. Хендерсона [31], другие описаны в статье [21], здесь показаны лишь те, что нужны для иллюстрации взаимодействия потоков и работы с общей памятью.

$M(SECD)^+$ – обозначение многопроцессорного комплекса над общей памятью, оно символизирует, что M – общая память для всех процессоров, а $(SECD)^+$ – что хотя бы один процессор обязателен, общее число процессоров произвольно, и не исключено изменение их числа в динамике.

При описании шагов процесса используются следующие обозначения:
 s – Stack – стек для окончательных и промежуточных результатов,
 e – Environment – контекст для размещения локальных значений переменных,
 c – Control_list – управляющая вычислениями программа,
 d – Dump – резервная память для обеспечения надёжности вычислений.
 m – Memory – общая память, выглядит как внешний регистр, доступный всем процессам.

$m s e c d \rightarrow m' s' e' c' d'$ – шаг процесса

– переход от старого состояния к новому.

$m s e c d \rightarrow \{ m' s' e' c' d' \mid \dots \}$

– переход, изменяющий состояние группы процессов.

$\{ m s e c d \mid \dots \} \rightarrow m' s' e' c' d'$

– переход, зависящий от группы процессов.

$AM = m s e c d$

– именование текущего состояния процесса на одном процессоре.

AM_k – имя отдельного процесса AM с номером k .

– Процесс знает свой номер « k ».

Шаги работы виртуальной машины происходят как изменение состояния процесса или изменение состояний группы процессов. Возможны переходы, требующие учёта состояний группы процессов.

Для четкого отделения обрабатываемых данных от остальной части списка в описания команд используются обозначения, как в книге П. Хендерсона:

($x . L$) – это список, в котором первый, головной элемент списка – x , а остальные, хвостовые находятся в списке L ;

($x y . L$) – в таком списке первый элемент – x , второй элемент – y , остальные находятся в списке L и т. д.

Для поддержки параллельных вычислений требуется несколько команд по организации комплексов процессов (неупорядоченных наборов – KIT) и рядов действий (последовательностей – ROW) с передачей их результатов и явной обработкой ошибок. Процессы образуются следующими командами:

KIT – комплекс из процессов для выполнения рядов действий, построенных в результате компиляции комплекта потоков;

ROW – ряд действий, созданный при компиляции потока как один процесс;

REZ – результат процесса и его номер размещаются в стеке;

WAIT – ожидание приостановки процесса (завершения или сообщения);

SEND – сообщение и его тэг передаются указанному процессу;

NEXT – ожидание полного завершения предыдущего действия процесса.

$$\begin{aligned} AM0 &= m (2^1 . s) e (KIT c1 c2 . c) d \\ &\rightarrow \{ AM0 = m s e c d \\ &\quad | AM1 = @m s e c1 d \\ &\quad | AM2 = @m s e c2 d \\ &\quad \} \end{aligned}$$

Порядок порождения процессов произвольный. Процесс $AM0$ работает как монитор на основном процессоре и существует по умолчанию. В данном примере в произвольном порядке создано два процесса $AM1$ и $AM2$, для каждого из которых порождён свой процессор. Общая память для порождённых процессоров доступна через ссылку.

Введено обозначение:

@ x – адрес позиции « x » или доступ к общей памяти в целом из процессора.

Процесс выполняется как ряд действий. Виртуальная машина поочерёдно

1 Для простоты иллюстраций рассматривается два процесса, а в процессе два действия.

запускает действия ряда, выделяя каждому из них свой процессор. Запуски действий происходят в порядке записи в программе, а завершения действий могут происходить в произвольном порядке, что соответствует множеству допустимых процессов, показанному на примерах раздела 1.

$$\begin{aligned} &AMt = m (2 . s) e (ROW c1 c2 . c) d \\ &\rightarrow \{ AMt = m (1 . s) e (ROW c2 . c) d \\ &\quad | AM1 = @m (1 . s) e (c1 SEND 1 AMt) d \\ &\quad \} \end{aligned}$$

Процесс AMt запускает из двух предстоящих действий первое – c1, создавая для него свой процесс AM1, и уменьшает счётчик предстоящих действий ряда на 1. В стеке процесса стоит его номер. Созданный процесс AM1 сообщит процессу AMt о завершении выполнения действия c1.

$$\begin{aligned} &AMt = m (1 . s) e (ROW c2 . c) d \\ &\rightarrow \{ AMt = m s e c d \\ &\quad | AM2 = @m (2 . s) e (c2 (SEND 2 AMt)) d \\ &\quad \} \end{aligned}$$

Потом, когда-нибудь, процесс AMt запустит второе действие – c2 и для него создаст другой процесс AM2. Уже оставался 1 элемент – выход из рекурсии порождения процессов ² для выполнения ряда действий. Завершилось ли первое действие – пока не известно. Если нужна строгая очерёдность действий, её надо при компиляции обеспечивать сообщениями или командой NEXT.

По мере завершения ряда действий процесса выполняется сборка результата командой REZ, выполняющей запись в стек число результатов комплекса процессов или результаты действий с номерами потоков.

$$\begin{aligned} &\{ AMt = m s e c d \quad \% c2 \text{ уже запущен,} \\ &\quad | AM1 = @m (a 1 . s) e (REZ) d \end{aligned}$$

² Здесь в процессе два действия

```
% получен результат c1 для передачи в AMt, возможно
% требующий свёртки для строгого результата.
| AM2 = @m (2 . s) e (WAIT AM1 c2) d
% c2 дожидается завершения c1
}% переход по готовности трёх процессов
```

```
→ { AMt = m (1 a . s) e (ass 2. c) d
    % AMt забирает результат c1 и
    % ждёт свёртки с результатом c2.
    | AM2 = @m (2 . s) e c2 d
    % c2 становится исполнимым фрагментом
    }
```

ass – представление функции, выполняющей свёртку указанного числа элементов в стеке и выдающей одно данное, рассматриваемое как строгий результат. Например: список, структура, сумма, произведение, макс, мин, последний и т. п. Со временем и второй процесс будет завершён, что позволит основному процессу выполнить свёртку результатов этих двух процессов.

6. НЕИЗМЕНЯЕМОСТЬ И/ИЛИ ВОССТАНОВЛЕНИЕ ДАННЫХ

Локальные переменные, хранимые в памяти обычных процессов, подчинены принципу неизменяемости, а глобальные переменные в общей памяти доступны через блок прямого доступа, для них вместо неизменяемости поддержан механизм восстановления отеснённых присваиваниями значений. Наличие структуро сохраняющих чисто функциональных эквивалентов для деструктивных функций гарантирует гладкий спуск от строгого функционального программирования в направлении дозированной императивности. Для возможности восстанавливать значения переменных они сопровождаются протоколом, в котором новые значения размещаются вслед за именем переменной, а к прежним данным можно вернуться при необходимости, например, при отладке посмотреть историю изменения данных.

При выполнении обратимых действий допускается и обратимость воздействий на общую память, хранящую переменные с неизменяемыми указателями

на изменяемое значение. Каждый элемент общей памяти в любой момент времени обрабатывается только в одном процессе программы, подобно захвату-освобождению файла или устройства. В протоколе изменений можно видеть какой процесс изменил значение данной переменной и, при необходимости, вернуть утраченное значение.

Для описания команд работы с общей памятью используются дополнительные обозначения:

[n] – содержимое n-го элемента общей памяти.

r=[...] – установка значения регистра виртуальной машины.

m=[. . . x ... y . . .] – можно указывать отдельные элементы памяти.

[... Var: Val ...] – размеченное множество, содержащее элемент Val, помеченный меткой Var.

Регистр «m» представляет собой размеченное множество, метки в котором являются именами переменных и допускают прямой доступ, а помеченные элементы расположены в куче, содержат значения переменных и протоколы их изменения, выглядящие как последовательность имён процессов с установленными этими процессами значениями.

Если память «m» содержит для переменной X хранимое в памяти значение xt, установленное процессом AMt, то хранящийся в памяти элемент имеет вид «X: (xt AMt xt AM1 x1 AM2 x2 . . .)», где помеченный именем переменной X список содержит её текущее значение xt, а далее размещены имена процессов и ранее установленные этими процессами значения переменной. Здесь показано, что переменная X ранее имела значения x1 и x2, установленные процессами AM1 и AM2 соответственно, а теперь процесс AMt установил ей значение xt. Если после этого процесс AMk присвоит переменной X значение xk, то после такого изменения элемент памяти приобретёт вид

m'=[. . . X : (xk AMk xk AMt xt AM1 x1 AM2 x2 . . .)].

Произойдёт побочный эффект присваивания, допускающий при необходимости восстановление прежних значений.

Воздействия на общую память сводятся к пересылкам и обменам данными:

LD – размещение в стек «s» константы из регистра управления «C».

MLL – пересылка головного элемента из одного списка в другой;

MLV– пересылка головного элемента списка в элемент вектора;

CHNG – обмен данными в общей памяти комплекса.

При пересылке элемент из списка исчезает, а появляется в другой структуре данных. Команды пересылок и обмена данными в общей памяти рассматриваются как средства низкого уровня для профилактики возникновения временных интервалов между парами взаимосвязанных воздействий на память на уровне аппаратуры.

Для примера рассмотрим подробнее команду CHNG – обмен данными в общей памяти. При компиляции выражения « $X \Leftrightarrow Y$ », представляющего в потоке R обмен значениями переменных X и Y, где $X=x$ и $Y=y$, в скомпилированном для него коде процесса AR будет выстроен список команд вида «(LD Y CHNG X . c)». Это значит, что сначала произойдёт загрузка имени переменной Y в стек текущего процесса, затем будет обмен значениями с переменной X из регистра управления. Обмен значениями происходит в общей памяти, если в ней уже расположены значения этих переменных. Получается, что при организации обмена компилятор размещает в стеке имя одной переменной, а в регистре управления программой – имя другой переменной.

$$\begin{aligned} AR = m \ s \ e \ (\ LD \ Y \ CHNG \ X \ . \ c \) \ d \\ \rightarrow m \ (\ Y \ . \ s \) \ e \ (\ CHNG \ X \ . \ c \) \ d \end{aligned}$$

% Сначала команда LD подготовит первого участника

Пусть выполнена команда LD и память имеет состояние, содержащее значения переменных X и Y, а P_x и P_y – протоколы с прежними значениями этих переменных соответственно:

$$m = [\dots X : (x \ . \ P_x) \dots \quad Y : (y \ . \ P_y) \dots]$$

Это условие возможности выполнять команду обмена:

$$\begin{aligned} AR = m = [\dots X : (x \ . \ P_x) \dots \\ \quad Y : (y \ . \ P_y) \dots] \ (\ Y \ . \ s \) \ e \ (\ CHNG \ X \ . \ c \) \ d \\ \rightarrow m' = [\dots X : (y \ AR \ y \ . \ P_x) \dots \\ \quad Y : (x \ AR \ x \ . \ P_y) \dots] \ (\ (\ y \ x \) \ . \ s \) \ e \ c \ d \end{aligned}$$

Прежние протоколы P_x и P_y для переменных X и Y остались без изменений, но в новых протоколах будет ($X=y$, а $Y=x$) с указанием, что это сделал процесс AR.

Здесь процесс AR производит обмен значениями x и y между переменными X и Y , в результате обмена обе переменные получают новые значения, они в протоколах сопровождаются именем процесса AR, выполнявшего обмен, между действиями обмена стороннее вмешательство невозможно. Сформирован строгий результат из новых значений переменных X и $Y - (y\ x)$.

7. МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ОБЩЕЙ И ЛОКАЛЬНОЙ ПАМЯТИ

Освобождение общей памяти может использовать решения, подобные опробованным в практике операционных систем при организации управления распределёнными системами с совмещением работы независимых программ. Обычно такие решения используют специальные структуры данных. Здесь рассмотрим варианты использования паспорта процесса и выделения пассивного процессора общей памяти.

Допустим, что каждый локальный процесс вместо доступа к общей памяти обладает вектором с именами или адресами используемых в нём глобальных переменных, что можно рассматривать как паспорт процесса, работающий как размазанная переменная. При освобождении локальной памяти процесса формируется новое, возможно уменьшенное, состояние этого вектора, доступное в любой момент и процессору общей памяти. Обычные процессы вместо полного доступа к общей памяти « m » в таком случае обладают лишь вектором « v » для доступа к определённым регистрам из блока прямого доступа. Вместо формата « $m\ s\ e\ c\ d$ » тогда используется формат « $v(m)\ s\ e\ c\ d$ », где « $v(m)$ » – регистр для паспорта процесса, использующего общую память « m ».

$v(m)\ s\ e\ c\ d \rightarrow v'(m')\ s'\ e'\ c'\ d'$ – переход от старого состояния виртуальной машины к новому, где « v » – вектор доступа к глобальным переменным из общей памяти « m ».

Обычные процессы несколько изменяют метод взаимодействия с общей памятью. Она уже не принадлежит им равноправно, а доступна через паспорт – вектор указателей на данные в общей памяти. Различие можно показать на примере команды MLL – пересылка из головного элемента одного списка в другой список, где этот элемент становится головным. Списки доступны через указатели x и y для переменных X и Y соответственно.

Пусть паспорт процесса AM содержит указатели на значения двух переменных X и Y. Это значит, что в общей памяти «m» сохраняются соответствующие этим переменным списки со значениями x и y, ранее установленные возможно другими процессами.

```
v (m)=[X: @x Y: @y ...]
m=[X: @x ... Y: @y ... X: (x AMi (xh . xt) . Px) ...
      Y: (y AMj (yh . yt) . Py)
  ... ]
```

По указателям на значения этих переменных x и y в памяти M доступны данные «x=(xh . xt)» и «y=(yh . yt)»:

```
m=[x=(xh . xt) y=(yh . yt)
  % значения переменных X и Y до обмена
  X: @x Y: @y          % паспорт процесса AM
X: (x AMi (xh . xt) . Px) Y: (y AMj (yh . yt) . Py)
  % данные переменных X и Y
  ... ]
```

Новое состояние памяти должно приобрести вид

```
m'=[x=xt y=(xh yh . yt)   % значения после обмена
  X: @x Y: @y          % паспорт без изменений
  X: ( x AM x AMi (xh. xt) . Px )
  Y: ( y AM y AMj (yh . yt) . Py)
  % данные о переменных изменены
  % при неизменном указателе
  % указан AM, изменивший значения по указателю
  ...]
```

В таком случае возможно выполнение команды MML для переменных X и Y.
 $v(m) (X . s) e (MLL Y . c) d \rightarrow v(m') (X Y . s) e c d$

Или более подробно:

$$\begin{aligned} &v(m=[x=(xh . xt) y=(yh . yt) X: @x Y: @y \\ &X: (x AM_i (xh . xt) . P_x) \\ &Y: (y AM_j (yh . yt) . P_y] \dots) (X . s) e (MLL Y . c) d \\ \rightarrow &v(m'=[x= xt y=(xh yh . yt) X: @x Y: @y \\ &X: (x AM x AM_i (xh . xt) . P_x) \\ &Y: (y AM y AM_j (yh . yt) . P_y \\ &\dots]) (X Y . s) e c d \end{aligned}$$

Изменилось состояние памяти «m», доступное по прежним адресам изменённых значений x и y , и пополнились протоколы оттеснённых значений переменных X и Y . Паспорт не изменился, но он теперь ссылается на изменённое состояние общей памяти.

Такой подход при освобождении общей памяти даёт возможность частичного освобождения памяти, не задействованной в объединении паспортов локальных процессов, без приостановки всех процессоров.

Несколько проще модель со специальным процессором общей памяти, поддерживающим пассивный процесс, память которого размечена на занятые и свободные регистры, а команды выполняются лишь по запросам из активных процессов. В таком случае многопроцессорный комплекс состоит из обычных процессоров и одного процессора общей памяти, с которым могут взаимодействовать обычные процессоры. Между собой процессоры взаимодействуют только через общую память. Каждый обычный процессор может выполнять активный процесс из ряда команд, среди которых встречаются команды запросов к пассивному процессу общей памяти. Это обычные команды, выполняемые по ходу активного процесса без особенностей, управляющие доступом к общей памяти, всеми обменами данных и пересылками значений переменных, упомянутыми в разделе 5.

Команды запроса активных процессов к общей памяти имеют двойников среди команд пассивного процесса общей памяти. Это команды типа ленивых вычислений, возбуждаемые при выполнении запросов из активных процессов. Ко-

манды пассивного процесса общей памяти структурированы в ряд очередей, каждая из которых соответствует активному процессу и содержит двойников запросов в том же порядке, что и в активном процессе. Пара запрос из активного процесса и его двойник из пассивного процесса здесь названа «дублет». Она выполняются неразделимо, одновременно в стиле рандеву языка Ada. Механизм рандеву обеспечивает исключение случайного вмешательства сторонних процессов в работу общей памяти. (см. табл. 3).

Таблица 3. Дублеты – парные команды – срабатывают только неразрывно вместе

<i>Запросы</i>	<i>Двойники запросов</i>	<i>Неразделимая пара</i>	<i>Примечание</i>
GIVE	TAKE	GIVE-TAKE	Запрос на переменную
SAFE	WRITE	SAFE-WRITE	Запись значения переменной
READ	VAR	READ-VAR	Запрос значения переменной
UNDO	UNDO	UNDO-UNDO	Восстановить прежнее значение переменной
FREE	FREE	FREE-FREE	Освободить память от переменной

N – Имя переменной для данного в общей памяти. Все имена различны, они известны во всех процессах, определены при компиляции.

@*N* – адрес текущего значения переменной *N* в регистре, хранящем и протокол изменения значений – историю присваиваний.

<@*N* ... > – шкала свободных регистров, содержащая адрес @*N*.

K – кратность доступа к переменной в текущий момент – число локальных процессов.

*N***K* – имя переменной с указанным числом использующих её процессов.

((*v s e* (*C1 . c1*) *d* | *v s e* *C2 . c2*) *d*)) – одновременное срабатывание первых команд *C1* и *C2* из регистров «с» двух процессов.

Для примера рассмотрим команду-дублет GIVE-TAKE. При её выполнении на стеке процесса s имя переменной H замещается на адрес её значения $@H$, а имя переменной заносится в паспорт процесса v . Одновременно $@H$ адрес в общей памяти исчезает из шкалы свободных регистров.

$((v (H . s) e (GIVE . c) d \rightarrow (H . v) (@H . s) e c d$

% адрес для значения в стеке

$| m=[<@H \dots > \dots] s e (TAKE . c) d \rightarrow m'=[\dots H*1 \dots] s e c d$

% новая переменная, счётчик был установлен в 0, теперь он = 1.

$)$ *% команды GIVE и TAKE срабатывают только одновременно*

% переменная H уже была,

% ей уже выделена память другим процессом.

$((v (H . s) e (GIVE . c) d \rightarrow (H . v) (@H . s) e c d$

$| m=[\dots (H*K): Ph \dots] s e (TAKE . c) d$

$\rightarrow m'=[\dots (H*(K+1)): Ph \dots] s e c d$

$)$

% команды GIVE и TAKE срабатывают одновременно

Описанная схема работы с памятью позволяет объединять методы освобождения памяти типа «сборки мусора» с методами счётчиков доступа при условии исключения слияния ссылок.

ЗАКЛЮЧЕНИЕ

В статье описаны решения, принятые в проекте разрабатываемого в Лаборатории информационных систем ИСИ СО РАН учебного языка программирования СИНХРО, предназначенного для ознакомления с параллелизмом и особенностями многопоточного программирования, а также приобретения навыков организации взаимодействия процессов над общей памятью и профилактики страха перед побочными эффектами при определении сценариев отладки учебных программ. Главные из них – поддержка независимости порядка вычислений от порядка записи выражений в программе в соответствии с принципом независимости параметров, расширение понятия «данное» на представление процессоров и процессов, а понятия «структуры данных» – на сети взаимодействующих потоков

в соответствии с принципом универсальности, учёт динамики критериев при переходе от одной парадигмы программирования к другой в соответствии с принципом гибкости ограничений, описание системы команд виртуальной машины для определения учебного языка параллельного программирования в соответствии с принципом само-применимости. Кроме того, предложены две модели работы с общей памятью, позволяющие от неизменяемости данных перейти к методам восстановления данных. Такие решения смягчают противопоставление методов строго функционального программирования, нацеленного на правильность программ, и императивно-процедурного программирования, позволяющего привычными способами повышать эффективность и производительность программ.

Рассматривая задачу использования учебных языков параллельного программирования как путь к решению проблемы адаптации методов вычислений к различным особенностям используемых многопроцессорных комплексов и многоядерных процессоров, следует видеть, что решение этой проблемы, кроме ознакомления с проблемами и методами, требует разработки новых методик обучения и методов реализации программ с акцентом на тестирование, верификацию и отладку, а также развития средств и методов ясного описания семантики языков параллельного программирования, включая средства метапрограммирования для представления программируемых преобразований текста и переносимого кода программы с удостоверением их корректности.

Предлагаемый язык СИНХРО представляет собой эксперимент по выбору базовых средств для достаточно полного решения проблем эффективной реализации параллельных алгоритмов, вынужденно требующих использовать весьма широкий спектр сложно совместимых средств: от управляющих действий более низкого уровня, чем в привычных языках программирования, до манипулирования пространствами решений по обработке данных в памяти, типичных для языков сверхвысокого уровня. Обзор исследований в смежных областях представлен в работах [32, 33].

Особый круг образовательных проблем связан с навыками учёта особенностей многоуровневой памяти в многопроцессорных системах. Обычное программирование такие проблемы может не замечать, полагаясь на решения компилятора, располагающего статической информацией о типах используемых данных и способного при необходимости выполнить оптимизирующие преобразования

программы. Новые и долгоживущие языки программирования, как правило, имеют мультипарадигмальный характер, что приводит к идее их расширения на многие модели параллельных вычислений.

Новые языки программирования содержат разнообразные конструкции для представления программ параллельных вычислений. Нужна система обучения, приспособленная не только к ознакомлению с отдельными эффектами и средствами параллельных вычислений, но и поддерживающая осознание особенностей взаимодействия процессов, удобная для экспериментов по формированию интуитивных навыков подготовки работоспособных программ и понимания новых возможностей аппаратуры. Не исключено, что этому противостоит исторически сложившаяся традиция одномерного, линейного представления текстов программ в виде бесконечной строки. В данной статье рассмотрен ряд форм, которые могут быть полезны при решении учебных задач по организации взаимодействующих параллельных процессов, включая процессы над общей памятью, на уровне решений, предложенных в языке учебного программирования СИНХРО.

Во введении определены цели опережающего ознакомления с параллелизмом. В нестрогом описании основных механизмов языка СИНХРО отмечена его структура как череда диалектов и расширенная трактовка ряда понятий программирования (раздел 1). Сформулирован ряд решений по созданию языка СИНХРО, включая учёт скрытой грамматики деятельности в процессе обучения, эксперимент по относительному описанию диалектов языка программирования и гибридных механизмов взаимодействия локальной и общей памяти (раздел 2). При описании ряда особенностей семантики языка особо показано различие в мощности семейства допустимых процессов вычисления для функционально эквивалентных семантически различных конструкций (раздел 3). Показана прагматика перехода от потоков к процессам, включая вытекающие из этого решения по системе команд, обеспечивающие неделимость действий по обмену данными (раздел 4), а также, дополняющие неизменяемость данных возможностью их восстановления (раздел 5) и допускающие механизм взаимодействия разных методов освобождения памяти (раздел 6).

Благодарность. Благодарю Дмитрия Владимировича Мажугу за экспериментальную реализацию ядра языка СИНХРО на языке функционального программирования Clojure.

СПИСОК ЛИТЕРАТУРЫ

1. *Городняя Л.В.* Язык параллельного программирования Синхро, предназначенный для обучения. Новосибирск: ИСИ им. А.П. Ершова СО РАН, 2016. 30 с. (Препринт/ИСИ СО РАН; № 180).

URL: https://www.iis.nsk.su/files/preprints/gorodnyaya_180.pdf.

2. *Городняя Л.В.* О курсе «Начала параллелизма» // Ершовская конференция по информатике. Секция «Информатика образования». 27 июня 2011 года. Новосибирск: ИСИ им. А.П. Ершова СО РАН. С. 51–54.

3. *Пеппер П., Экснер Ю., Зюдхольд М.* Функциональный поход к разработке программ с развитым параллелизмом // Системная информатика. Вып 4. Методы теоретического и системного программирования. Новосибирск: Наука. Сиб. изд. фирма, 1995. С. 334–360.

4. *Воеводин В.В., Воеводин Вл.В.* Параллельные вычисления. СПб.: БХВ-Петербург, 2002. 608 с.

5. *Хоар Ч.* Взаимодействующие последовательные процессы. М.: Мир, 1989. 264 с.

6. *Быстров А.В.* Сетевые средства синхронизации процессов // Тр. Всесоюзного научно-технического семинара «Программное обеспечение многопроцессорных систем». Калинин, 1985. С. 44–46.

7. *Городняя Л.В.* Парадигмы программирования. Часть 4. Параллельное программирование. Новосибирск, 2015. 74 с. (Препр. / ИСИ СО РАН; № 175). URL: https://www.iis.nsk.su/files/preprints/gorodnyaya_175.pdf.

8. *Лавров С.С.* Расширяемость языков. Подходы и практика. В сб.: Прикладная информатика, вып. 2. М.: Финансы и статистика, 1984. С. 17–22.

9. *Городняя Л.В.* О неявной мультипарадигмальности параллельного программирования. Материалы конференции «Научный сервис в сети Интернет: труды XXIII Всероссийской научной конференции». С. 104–116.

URL: https://library.keldysh.ru/prep_vw.asp?pid=9203.

10. *Городняя Л.В.* Методика парадигмального анализа языков и систем программирования // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2019. С. 262–277.

URL: <http://keldysh.ru/abrau/2019/theses/03.pdf> doi:10.20948/abrau-2019-03.

11. *Андреева Т.А.* и др. Компьютерные языки как форма и средство представления, порождения и анализа научных и профессиональных знаний // Тр. XV Всерос. научно-методической конф. «Телематика–2008». СПб., 2008. С. 77–78.

12. *Городняя Л.В.* Парадигмы программирования. Ч. 5. Учебные языки программирования. Новосибирск, 2015. 60 с. (Препр. / ИСИ СО РАН; № 176).

URL: http://www.iis.nsk.su/files/preprints/gorodnyaya_176.pdf.

13. *Звенигородский Г.А.* Первые уроки программирования. Библиотечка «Кванта». М.: Наука, 1985. Т. 41. URL: <http://cip.iis.nsk.su/files/course/zven-ves.pdf>.

14. *Кушниренко А.Г., Грибанова И.Н., Райко М.В., Зайдельман Я.Н.* Как мы проводим вводное занятие по алгоритмике в разновозрастной группе дошкольников и младших школьников. URL:

<https://fgoskomplekt.ru/upload/iblock/a2c/4o0ce5xyjyrha1b8qwsghlnjqp1iqme.pdf>.

15. Что такое Лого. URL: <http://www.int-edu.ru/logo/logo.html>.

16. *Голиков Д.В., Голиков А.Д.* Книга юных программистов на Scratch – SmashWords, 2013. ISBN 978-1310227554.

17. *Ершов А.П.* Программирование – вторая грамотность. URL: http://ershov.iis.nsk.su/ru/second_literacy/article.

18. *Брукс Ф.* Мифический человеко-месяц, или Как создаются программные системы. СПб.: Питер, 2021. 368 с.

19. *Городняя Л.В.* О функциональном программировании // Компьютерные инструменты в образовании (в печати).

20. *Городняя Л.В.* О проблеме автоматизации параллельного программирования // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров: Труды Международной суперкомпьютерной конференции (22–27 сентября 2014 г., г. Новороссийск). М.: Изд-во МГУ, 2014. С. 191–196.

URL: <http://agora.guru.ru/abrau2014>

21. *Городняя Л.В.* Абстрактная машина языка программирования учебного назначения СИНХРО // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2021. № 4. С. 16–35.

22. *Адамович А.И., Климов Анд.В.* Об опыте использования среды мета-программирования Eclipse/TMF для конструирования специализированных языков // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19–24 сентября 2016 г. г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2016. С. 3–8. URL: <https://doi.org/10.20948/abrau-2016-45>.

23. *Городняя Л.В.* Подходы к представлению синтаксиса языков программирования. Новосибирск, 2019. 39 с. (Препр. / ИСИ СО РАН; № 185). URL: https://www.iis.nsk.su/files/preprints/Preprint_185.pdf

24. About SETL and GNU SETL. URL: <https://setl.org/doc/setl.html>.

25. *Schwartz Jacob T.* Set Theory as a Language for Program Specification and Programming. Courant Institute of Mathematical Sciences, New York University, 1970.

26. *Kondratyev D.A., Promsky A.V.* Developing a self-applicable verification system. Theory and practice // Automatic Control and Computer Sciences. 2015. Vol. 49. Issue 7. P. 445–452. URL: <https://doi.org/10.3103/S0146411615070123>

27. *Малышкин В.Э.* Технология фрагментированного программирования // Параллельные вычислительные технологии (ПаВТ'2012): труды международной научной конференции (Новосибирск, 26–30 марта 2012 г.). Челябинск: Издательский центр ЮУрГУ, 2012. С. 598–605. URL: <http://omega.sp.susu.ru/books/conference/PaVT2012/short/212.pdf>

28. *Городняя Л.В.* Модели работы с памятью в учебном языке программирования СИНХРО // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции (19–22 сентября 2022 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2022. С. 137–154.

29. *Городняя Л.В.* Работа с данными в учебном языке программирования СИНХРО // Суперкомпьютерные дни в России: Труды международной конференции. 26–27 сентября 2022 г., Москва / Под. ред. Вл. В. Воеводина. М.: МАКС Пресс, 2022. ISBN 978-5-317-06875-2 e-ISBN 978-5-317-06876-9 <https://doi.org/10.29003/m3109.RussianSCDays2022>. С. 87–97. <https://doi.org/10.29003/m3109.RussianSCDays2022>(внешняя ссылка)

30. *McCarthy J.* LISP 1.5 Programming Manual. The MIT Press., Cambridge,

1963. 106 p.

31. Хендерсон П. Функциональное программирование. М.: Мир, 1983. 349 с.

32. Городняя Л.В. Гуманитарные факторы программирования /Сиб. Отделение Рос. Акад. наук, Ин-т систем информатики им. А.П. Ершова. Новосибирск: Изд-во СО РАН, 2020. 163 с.

33. Городняя Л.В. Функциональное программирование. Парадигма, модели, методы / Сиб. Отделение Рос. Акад. наук, Ин-т систем информатики им. А.П. Ершова. Новосибирск: Изд-во СО РАН, 2022. 482 с.

ORGANIZATION OF CALCULATIONS AND WORK WITH MEMORY IN THE EDUCATIONAL PROGRAMMING LANGUAGE SYNHRO

L. V. Gorodnyaya^{1, 2} [0000-0002-4639-9032]

¹ *A.P. Ershov Institute of Informatics Systems SB RAS, 6, Acad. Lavrentjev pr., Novosibirsk 630090, Russia*

² *Novosibirsk State University, 1, st. Pirogova, Novosibirsk, 630090, Russia*

Abstract

The article is devoted to a number of decisions made in the project of the educational programming language Synchro, which is being developed at the Laboratory of Information Systems of the *A.P. Ershov IIS SB RAS*, designed to familiarize with the basic phenomena of the interaction of processes and control of calculations over shared memory. The focus is on the functional programming paradigm. The language is aimed at schoolchildren of primary and secondary grades, as well as junior students and non-professionals. During training, the experience of operating with toy robots moving on a checkered board is used. The article is of interest to everyone who is interested in the problems of modern computer science, programming and information technology, especially the problems of parallel computing on supercomputers and distributed systems, and in general the use of multiprocessor systems.

Keywords: educational programming languages, virtual machine, command system, functional programming, data recovery, memory release, multithreaded programs, parallel computing, shared memory, process interaction.

REFERENCES

1. Gorodnyaya L.V. YAzyk parallelnogo programmirovaniya Sinkhro, prednaznachenny dlya obucheniya. Novosibirsk: ISI im. A.P. Yershova SO RAN, 2016. 30 s. (Preprint/ISI SO RAN; № 180).

URL: https://www.iis.nsk.su/files/preprints/gorodnyaya_180.pdf.

2. Gorodnyaya L.V. O kurse «Nachala parallelizma» // Yershovskaya konferentsiya po informatike. Sektsiya «Informatika obrazovaniya». 27 iyunya 2011 goda. Novosibirsk: ISI im. A.P. Yershova SO RAN. S. 51–54.

3. Pepper P., Eksner YU., Zyudkhol'd M. Funktsional'nyy pokhod k razrabotke programm s razvitym parallelizmom // Sistemnaya informatika. Vyp 4. Metody teoreticheskogo i sistemnogo programmirovaniya. Novosibirsk: Nauka. Sib. izd. firma, 1995. S. 334–360.

4. Voyevodin V.V., Voyevodin V.I. Parallelnyye vychisleniya. SPb.: BKHV-Peterburg, 2002. 608 s.

5. Khoar CH. Vzaimodeystvuyushchiye posledovatel'nyye protsessy. M.: Mir, 1989. 264 s.

6. Bystrov A.V. Setevyye sredstva sinkhronizatsii protsessov // Tr. Vsesoyuznogo nauchno-tekhnicheskogo seminar "Programmnoye obespecheniye mnogo-protsessornykh sistem". Kalinin, 1985. S. 44–46.

7. Gorodnyaya L.V. Paradigmy programmirovaniya. Chast' 4. Parallelnoye programmirovaniye. Novosibirsk, 2015. 74 s. (Prepr. / ISI SO RAN; № 175).

URL: https://www.iis.nsk.su/files/preprints/gorodnyaya_175.pdf.

8. Lavrov S.S. Rasshiryayemost' yazykov. Podkhody i praktika. V sb.: Prikladnaya informatika, vyp. 2. M.: Finansy i statistika, 1984. S. 17–22.

9. Gorodnyaya L.V. O neyavnoy mul'tiparadigmal'nosti parallelnogo programmirovaniya Material konferentsii "Nauchnyy servis v seti Internet: trudy XXIII Vserossiyskoy nauchnoy konferentsii. S. 104–116.

URL: https://library.keldysh.ru/prep_vw.asp?pid=9203.

10. *Gorodnyaya L.V.* Metodika paradigmal'nogo analiza yazykov i sistem programmirovaniya // Nauchnyy servis v seti Internet: trudy XXI Vserossiyskoy nauchnoy konferentsii (23–28 sentyabrya 2019 g., g. Novorossiysk). M.: IPM im. M.V. Keldysha, 2019. S. 262–277. URL: <http://keldysh.ru/abrau/2019/theses/03.pdf> URL: <https://doi.org/10.20948/abrau-2019-03>.

11. *Andreyeva T.A. i dr.* Komp'yuternyye yazyki kak forma i sredstvo predstavleniya, porozhdeniya i analiza nauchnykh i professional'nykh znaniy // Tr. XV Vseros. nauchno-metodicheskoy konf. «Telematika-2008». SPb., 2008. S. 77–78.

12. *Gorodnyaya L.V.* Paradigmy programmirovaniya. CH. 5. Uchebnyye yazyki programmirovaniya. Novosibirsk, 2015. 60 s. (Prepr. / ISI SO RAN; № 176). URL: http://www.iis.nsk.su/files/preprints/gorodnyaya_176.pdf.

13. *Zvenigorodskiy G.A.* Pervyye uroki programmirovaniya. Bibliotekha «Kvanta». M.: Nauka, 1985. T. 41. URL: <http://cip.iis.nsk.su/files/course/zven-ves.pdf>.

14. *Kushnirenko A.G., Gribanova I.N., Rayko M.V., Zaydel'man Ya.N.* Kak my provodim vvodnoye zanyatiye po algoritmike v raznovozrastnoy gruppe doshkol'nikov i mladshikh shkol'nikov. URL: <https://fgoskomplekt.ru/upload/iblock/a2c/4o0ce5xyjyrha1b8qwsghlnjqp1iqme.pdf>.

15. Chto takoye Logo. URL: <http://www.int-edu.ru/logo/logo.html>.

16. *Golikov D.V., Golikov A.D.* Kniga yunyx programmistov na Scratch – Smash-Words, 2013. ISBN 978-1310227554.

17. *Yershov A.P.* Programmirovaniye – vtoraya gramotnost'. URL: http://ershov.iis.nsk.su/ru/second_literacy/article.

18. *Bruks F.* Mificheskyy cheloveko-mesyats, ili Kak sozdayutsya programmnyye sistemy. SPb.: Piter, 2021. 368 s.

19. *Gorodnyaya L.V.* O funktsional'nom programmirovanii // Komp'yuternyye instrumenty v obrazovanii (v pechati).

20. *Gorodnyaya L.V.* O probleme avtomatizatsii parallel'nogo programmirovaniya // Nauchnyy servis v seti Internet: mnogoobraziye superkomp'yuternykh mirov: Trudy Mezhdunarodnoy superkomp'yuternoy konferentsii (22–27 sentyabrya 2014 g., g. Novorossiysk). M.: Izd-vo MGU, 2014. S. 191–196. URL: <http://agora.guru.ru/abrau2014>.

21. *Gorodnyaya L.V.* Abstraktnaya mashina yazyka programmirovaniya

uchebnogo naznacheniya SINKHRO // Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnyye tekhnologii. 2021. № 4. S. 16–35.

22. *Adamovich A.I., Klimov And.V.* Ob opyte ispol'zovaniya sredy metaprogrammirovaniya Eclipse/TMF dlya konstruirovaniya spetsializirovannykh yazykov // Nauchnyy servis v seti Internet: trudy XVIII Vserossiyskoy nauchnoy konferentsii (19–24 sentyabrya 2016 g. g. Novorossiysk). M.: IPM im. M.V. Keldysha, 2016. S. 3–8. URL: <https://doi.org/10.20948/abrau-2016-45>.

23. *Gorodnyaya L.V.* Podkhody k predstavleniyu sintaksisa yazykov programmirovaniya. Novosibirsk, 2019. 39 s. (Prepr. / ISI SO RAN; № 185). URL: https://www.iis.nsk.su/files/preprints/Preprint_185.pdf.

24. About SETL and GNU SETL. URL: <https://setl.org/doc/setl.html>.

25. *Schwartz Jacob T.* Set Theory as a Language for Program Specification and Programming. Courant Institute of Mathematical Sciences, New York University, 1970.

26. *Kondratyev D.A., Promsky A.V.* Developing a self-applicable verification system. Theory and practice // Automatic Control and Computer Sciences. 2015. Vol. 49. Issue 7. P. 445–452. URL: <https://doi.org/10.3103/S0146411615070123>.

27. *Malyshkin V.E.* Tekhnologiya fragmentirovannogo programmirovaniya // Parallel'nyye vychislitel'nyye tekhnologii (PaVT'2012): trudy mezhd.unarodnoy nauchnoy konferentsii (Novosibirsk, 26–30 marta 2012 g.). Chelyabinsk: Izdatel'skiy tsentr YUUrGU, 2012. S. 598–605. URL: <http://omega.sp.susu.ru/books/conference/PaVT2012/short/212.pdf>.

28. *Gorodskaya L.V.* Modeli raboty s pamyat'yu v uchebnom yazyke programmirovaniya SINKHRO // Nauchnyy servis v seti Internet: trudy XXIV Vserossiyskoy nauchnoy konferentsii (19–22 sentyabrya 2022 g., onlayn). M.: IPM im. M.V. Keldysha, 2022. S. 137–154.

29. *Gorodskaya L.V.* Rabota s dannymi v uchebnom yazyke programmirovaniya SINKHRO // Superkomp'yuternyye dni v Rossii: Trudy mezhdunarodnoy konferentsii. 26–27 sentyabrya 2022 g., Moskva / Pod. red. VI. V. Voyevodina. M.: MAKSS Press, 2022. ISBN 978-5-317-06875-2 e-ISBN 978-5-317-06876-9 URL: <https://doi.org/10.29003/m3109.RussianSCDays2022>. S. 87–97. URL: [https://doi.org/10.29003/m3109.RussianSCDays2022\(vneshnyaya_ssylnka\)](https://doi.org/10.29003/m3109.RussianSCDays2022(vneshnyaya_ssylnka)).

30. *McCarthy J.* LISP 1.5 Programming Manual. The MIT Press., Cambridge, 1963. 106 p.

31. *Khenderson P.* Funktsional'noye programmirovaniye. M.: Mir, 1983. 349 s.
 32. *Gorodnyaya L.V.* Gumanitarnyye faktory programmirovaniya / Sib. Otdeleniye Ros. Akad. Nauk, In-t sistem informatiki im. A.P. Yershova. Novosibirsk: Izd-vo SO RAN, 2020. 163 s.
 33. *Gorodnyaya L.V.* Funktsional'noye programmirovaniye. Paradigma, modeli, metody / Sib. Otdeleniye Ros. Akad. nauk, In-t sistem informatiki im. A.P. Yershova. Novosibirsk: Izd-vo SO RAN, 2022. 482 s.
-

СВЕДЕНИЯ ОБ АВТОРЕ



ГОРОДНЯЯ Лидия Васильевна – старший научный сотрудник Института систем информатики имени акад. Андрея Петровича Ершова СО РАН, доцент Новосибирского государственного университета, специалист в области системного программирования и образовательной информатики.

Lidia Vasiljevna GORODNYAYA – Senior Researcher of A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Associate Professor of Novosibirsk State University, a specialist in system programming and educational informatics.

email: lgorod@iis.nsk.su

ORCID: 0000-0002-4639-9032

Материал поступил в редакцию 10 декабря 2022 года

ИСПОЛЬЗОВАНИЕ ГРАФА СОАВТОРСТВА ДЛЯ ТЕМАТИЧЕСКОГО ПОИСКА КОНФЕРЕНЦИЙ ПО НАУКОМЕТРИЧЕСКИМ ДАННЫМ

А. С. Козицын¹ [0000-0002-8065-9061], С. А. Афонин² [0000-0003-3058-9269],

Д. А. Шачнев³ [0000-0002-5940-9180]

¹⁻³НИИ механики МГУ им. М.В. Ломоносова, Мичуринский пр., 1, Москва, 119192

¹alexanderkz@mail.ru, ²serg@msu.ru, ³mitya57@gmail.com

Аннотация

Применение современных методов тематического анализа для аналитической обработки больших объемов информации используется в настоящее время практически во всех сферах человеческой деятельности, в том числе, в наукометрии. Многие наукометрические системы и системы цитирования, включая всемирно известные WoS, Scopus, Google Scholar, разрабатывают тематические рубрики для поиска и обработки информации. Важными практическими задачами, которые могут решаться с применением методов тематической классификации, являются: оценка динамики развития тематических направлений в организации, в отдельной стране и мировой науке в целом; поиск статей по заданной тематике; поиск и оценка авторитетности экспертов; поиск журналов для публикации и другие актуальные задачи. Авторами созданы программные реализации алгоритмов для решения некоторых из перечисленных задач и ведутся научные исследования с целью создания новых эффективных математических моделей и алгоритмов в этой области.

Ключевые слова: тематический поиск, библиографические данные, поиск конференций, граф соавторства, информационные системы, наукометрия.

ВВЕДЕНИЕ

Тематический поиск информации используется в различных сферах деятельности. Например, тематический поиск экспертов [1] востребован в редакциях при поиске рецензентов, в том числе с учетом конфликтов интересов [2], на наукометрических производствах при поиске консультантов или потенциальных исполните-

лей проектов, в СМИ при поиске авторов для написания статей по заданной тематике. Тематический поиск журналов позволяет на основе аккумулярованного опыта научного сообщества автоматически подбирать журналы, соответствующие заданным научным интересам автора, и помогает молодым сотрудникам увеличить свою публикационную активность и повысить показатели цитируемости [3]. Применение шаблонов с условиями тематической принадлежности может использоваться для создания дополнительных логических правил обеспечения безопасности в информационных системах [4]. Применение средств тематического анализа для поиска конференций позволяет увеличивать полноту поиска и охват конференций, способствует расширению круга научного общения молодых ученых и формированию более тесных научных связей. Подобная функциональность позволяет значительно расширить возможности создаваемых в настоящий момент наукометрических систем для управления научными и образовательными организациями, улучшить качество оценки научной деятельности организации и стимулирования наиболее перспективных направлений научной деятельности [5].

Следует отметить, что осуществление тематического поиска, в том числе поиска конференций, требует использования специализированных поисковых информационных систем.

В настоящее время существует большое количество проектов, реализующих в той или иной степени хранение, рубрикацию и поиск конференций.

В интернете представлена значительная группа проектов, целью которых является продвижение услуг платных публикаций. Такие проекты, как правило, имеют небольшой охват конференций и не предоставляют сервисов качественного поиска. Например, в проекте konferencii.ru зарегистрировано в феврале 2022 года всего 7 конференций в области «Математика», причем все конференции «широкого профиля» (например, «Научный форум: технические и физико-математические науки»). Проект «Научные конференции России» (www.kon-ferenc.ru) также представляет конференции «широкого профиля» для быстрой публикации статей (например, «Диссеминация инновационного опыта как фактор модернизации науки и образования»). Проект «Научные-конференции.рф» (na-konferencii.ru) содержит около 30 конференций с датой начала в феврале 2022 и

не имеет тематического поиска. Проект «Социальная научная сеть» (www.science-community.org) содержит 12 конференций в области математики. Основным недостатком всех проектов этого типа является наличие большого процента фиктивных конференций, которые активно рекламируются для предоставления авторам возможности опубликовать свои работы на платной основе в максимально сжатые сроки (до трех дней). Такой «шум» не только затрудняет поиск, но и отпугивает организаторов авторитетных конференций, которые не хотят регистрировать свои мероприятия в подобном контексте, который можно охарактеризовать как «недобросовестный сегмент науки» [6].

Проект obshestvo.org специализируется на молодежных мероприятиях и содержит описание реальных конференций и олимпиад, но количество зарегистрированных в нем мероприятий также незначительно. За все время в рубрике «Математика» зарегистрировано 126 мероприятий.

Проект konferen.ru позиционируется как календарь научных конференций России и содержит около тысячи планируемых мероприятий: олимпиад; школ; конференций и других. Поиск на сайте возможен по словам из названия и по одной из 39 рубрик.

Проект поиска мероприятий WorldExpo (worldexpo.pro) специализируется на представлении коммерческих выставок, форумов, семинаров и, в меньшей степени, конференций.

В интернете представлены также зарубежные системы поиска конференций, например, проект conferencealerts.com предоставляет возможность поиска по рубрике и стране, проект www.allconferences.com охватывает около 50 тысяч конференций и предоставляет возможность фильтрации по стране, городу, названию и рубрике. Однако большая часть конференций, зарегистрированных в таких системах, носит коммерческий характер. Например, в системе www.allconferences.com зарегистрировано более тысячи конференций по сетевым протоколам, более шестисот конференций по электронной коммерции, около ста конференций по компьютерным играм и только десять конференций по алгоритмам. Кроме того, в таких системах, как правило, отсутствуют не только российские конферен-

ции, но и большинство международных конференций, проходящих на территории России. Например, по данным conferencealerts.com, в России в 2023 году будет проходить только 15 различных конференций.

Русскоязычные системы обзора конференций организаций (например, conf.msu.ru в МГУ им. М.В. Ломоносова, www.nstu.ru/science/scientific_events/search, НГТУ, www.ruthenia.ru, Тарусского университета) специализируются в большей степени на мероприятиях своей организации и имеют небольшой охват в масштабах России.

Существует отдельная категория систем, предназначенных для поддержки процесса организации конференций. Большая часть конференций, зарегистрированных в этих системах, имеет реально действующий программный комитет, рецензируемые доклады и, как следствие, хороший научный уровень. Например, система EasyChair (easychair.org), содержит большое количество конференций, но не имеет системы поиска. Ее российский аналог «Конференции России» (ruconf.ru) имеет систему контекстного поиска и поиска по классификатору, но количество загруженных в нее конференций незначительно (по машиностроению – 1, по медицине – 2 и так далее).

На основании приведенного обзора можно сделать вывод, что в России не существует хороших систем тематического поиска конференций, которые бы позволяли производить фильтрацию по конкретной тематике и учитывали бы при фильтрации авторитетность конференции.

СПОСОБЫ ТЕМАТИЧЕСКОГО ПОИСКА

В зависимости от данных, содержащихся в информационных системах, и информационной потребности пользователя [7] могут использоваться разные способы осуществления поиска. Самым простым является поиск по ключевым словам с использованием полных текстов докладов конференций. Методы полнотекстового поиска широко используются для подбора ресурсов в интернете как глобальными поисковыми системами (Google, Yandex), так и специализированными поисковыми системами, предназначенными для подбора материала из собственных баз данных. Метод прост в реализации, с использованием обратных индексов позволяет обеспечить высокую скорость поиска и поддерживается многими СУБД

[8–11]. Основным недостатком такого метода поиска является невозможность его проведения при отсутствии таких полнотекстовых данных. При регистрации библиографических данных в наукометрических системах авторы публикаций и докладов в обязательном порядке указывают название публикации или доклада, список соавторов и другие выходные данные работы, но не всегда размещают полный текст своей статьи или доклада. Это может быть обусловлено правовыми отношениями с издательствами, отсутствием текста доклада в нужном формате, нежеланием выполнять «лишние» действия в системе и другими причинами. Как следствие, в наукометрической системе может отсутствовать исходная информация для проведения полнотекстового поиска по ключевым словам.

Альтернативным способом поиска является поиск с использованием опорного множества авторов [12]. Опорное множество $H = \{(a, v)\}$, состоящее из пар «автор»–«вес автора», может вычисляться тремя различными способами: с использованием поиска экспертов [13] по ключевым словам; по выбранному пользователем журналу или по конференции. Вес авторов в опорном множестве может быть как действительным числом, так и элементом множества $\{0,1\}$. Во втором случае построение опорного множества значительно упрощается, однако снижается точность последующего поиска.

При использовании поиска экспертов пользователь должен указать множество ключевых слов, наилучшим образом описывающих его информационную потребность. По указанным ключевым словам поисковая система производит подбор экспертов с использованием алгоритмов [14], [13] или других и составляет взвешенное опорное множество H , которое будет использоваться для дальнейшего поиска. Дополнительно к ключевым словам для конкретизации информационной потребности могут указываться: тематические рубрики классификаторов, например, ГРНТИ; возможные места работы экспертов; возможность использования автоматического перевода ключевых слов при поиске и другие параметры.

Во втором случае для построения опорного множества авторов пользователь указывает образец для поиска в виде журнала или конференции, которые соответствуют выбранной им тематике. Поисковая система на основе указанного пользователем образца вычисляет взвешенное опорное множество на основе информации об авторах статей или докладов из выбранного источника. Для этого

для каждой публикации d каждому соавтору a ставится в соответствие определенный вес по формуле [15]

$$w(a, d) = \frac{1}{2} \delta_{1n} + \frac{1}{2k},$$

где n – порядковый номер соавтора в библиографическом списке публикации, а k – количество соавторов публикации a .

Множество весов авторов a вычисляется как максимальное значение его веса по всем публикациям журнала или конференции по формуле

$$v = \max_d (w(a, d)).$$

В опорное множество H включаются авторы, для которых $v > 0$.

Во всех трех случаях результатом работы первого шага алгоритма является построенное опорное множество $H = \{(a, v)\}$, содержащее список авторов, релевантных поисковому запросу пользователя.

На втором шаге алгоритма для всех авторов из построенного опорного множества рассматривается множество всех докладов, авторами которых они являются. Степень тематической близости конференции C запросу пользователя определяется на основании количества докладов, входящих в построенное множество, степени участия автора в докладе [16] и веса автора в опорном множестве:

$$F(c) = \sum_{(a,v) \in H} \min(v, \max_{d \in c} (w(a, d))).$$

Таким образом, степень соответствия конференции запросу определяется ее популярностью у других авторов, интересующихся соответствующей тематикой.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ


Апробация алгоритма производилась на данных наукометрической системы ИСТИНА [17, 18]. Объем данных представлен на странице статистики системы (Рис. 1).

Тип объекта	▲ Количество объектов
Дипломных работ	114362
Диссертаций	32871
Докладов на конференциях	350582
Докторов наук	10671
Достижений	21632
Кандидатов наук	33190
Книг	86533
Курсовых работ	53070
Наград	25556
Научных коллективов	622
Научных отчётов	10699
НИР	22514
Патентов	15351
Пользователей	170785
Пользователей из МГУ	28729
Пользователей, новых за последнюю неделю	850
Почетных членств в организациях	678
Преподаваний курсов	251710
Свидетельств о регистрации прав на ПО	2887
Соавторов	577380
Стажировок	23292
Статей	899377
Тезисов докладов	187984
Учебных курсов	122980
Членств в диссертационных советах	8429
Членств в научных обществах	9687

Рис. 1. Данные в системе ИСТИНА

Программная реализация трех описанных выше механизмов поиска выполнена на встроенном языке СУБД системы [19], для разработки интерфейса использовался механизм SQLReport [20].

При поиске по ключевым словам пользователю предлагается указать список ключевых слов, выбрать метод поиска, ограничить по желанию область поиска по рубрике ГРНТИ и подразделениям организации (Рис. 2).

 **ИСТИНА** Козицын Александр Сергеевич (sasha) [Выйти из системы](#)
Интеллектуальная Система Тематического Исследования НАукометрических данных
[Главная](#) [Для ответственных](#) [Моя страница](#) [Добавить работу](#) [Поиск](#) [Статистика](#) [О проекте](#) [Помощь](#)
[Администрирование](#)

i В связи с техническими работами в центре обработки данных, возможность загрузки и скачивания файлов временно недоступна. [Скрыть](#)

▲ Введите параметр Ключевые слова (через запятую) [Скрыть](#) [Скрыть](#)
[ВСЕ](#)

Ключевые слова (через запятую)

Область знаний

Подразделение

Встроенный поиск

Учет переводов

Рис. 2. Пример интерфейса поиска конференций с использованием поиска экспертов по ключевым словам

Результатом поиска является список конференций с указанием их соответствия запросу (Рис. 3).

При поиске конференций с использованием эталонного журнала или конференции пользователь сначала должен выбрать один журнал или одну конференцию, которые представляют для него интерес, и перейти по ссылке для поиска конференций, похожих по тематике.

Результаты поиска представляются в виде списка найденных конференций с указанием степени близости к эталонному объекту, а также с возможностью дальнейшего перемещения по графу близости конференций (Рис. 4).

Список конференций по теме с поиском экспертов

Show by 10 items Search:

V	NAME
0.65	XXVIII Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов 2021"
0.381	XXI Менделеевский съезд по общей и прикладной химии
0.294	Ломоносовские чтения - 2017
0.285	XX Менделеевский съезд по общей и прикладной химии
0.225	Ломоносовские чтения - 2016
0.192	VII Всероссийская Каргинская конференция "Полимеры-2017"
0.155	Научная конференция грантодержателей РФ «Фундаментальные химические исследования XXI-го века»
0.146	XIX Менделеевский съезд по общей и прикладной химии
0.128	VI Всероссийская Каргинская конференция "Полимеры-2014"

Рис. 3. Результат поиска с использованием механизмов подбора экспертов

Show by 10 items Search:

N	Конференция	Вес	Количество докладов	Похожие конференции	Похожие журналы
1	Ломоносовские чтения 2021. Секция вычислительная математика и кибернетика, 20-29 апреля 2021(2021)	3,94	136	конференции	журналы
2	ЛОМОНОСОВСКИЕ ЧТЕНИЯ Научная конференция Секция механики(2021)	3,25	155	конференции	журналы
3	XXVIII Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов 2021"(2021)	2,85	979	конференции	журналы
4	XII Международная научная конференция «Интеллектуальные системы и компьютерные науки»(2021)	1,87	86	конференции	журналы
5	27-я МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ по компьютерной лингвистике и интеллектуальным технологиям Диалог(2021)	1,67	1	конференции	журналы
6	Математические методы распознавания образов (ММРО-20)(2021)	1,58	9	конференции	журналы
7	XXII Международная конференция по вычислительной механике и современным прикладным программным системам (ВМСППС'2021)(2021)	1,58	11	конференции	журналы
8	CAICS 2020: NATIONAL CONGRESS ON COGNITIVE RESEARCH, ARTIFICIAL INTELLIGENCE AND NEUROINFORMATICS(2020)	1,54	19	конференции	журналы
9	VII Актуальные проблемы системной и программной инженерии (АПСПИ 2021)(2021)	1,45	5	конференции	журналы
	3rd International Science and Technology Conference «Modern				

Рис. 4. Поиск конференций по журналу

Для создания единого информационного пространства поиска граф близости конференций объединен с графом близости журналов [21]. В текущей реализации системы ИСТИНА доступ к поиску и навигации по графу близости журналов и конференций осуществляется из карточки журнала по ссылке «Похожие по тематике журналы». В дальнейшем планируется создать отдельный модуль работы с конференциями и добавить возможность поиска в карточку конференции.

Апробация разработанных авторами алгоритмов проводилась на данных наукометрической системы ИАС ИСТИНА по следующей методике. Программным реализациям каждому из трех алгоритмов было передано на вход пять запросов для поиска. В результатах запроса оценивались лучшие десять вариантов по следующей шкале: 2 – соответствует запросу; 1 – частично соответствует запросу; 0 – не соответствует запросу. Итоговая сумма баллов приводится в таблице 1.

Таблица 1. Результаты тестирования

Алгоритм	Точность, %
Поиск по журналу	81
Поиск по конференции	76
Использование поиска экспертов	89

Представленные результаты тестирования показывают, что наиболее точным способом поиска является использование алгоритма с тематическим поиском экспертов. Поиск по журналу обладает преимуществом перед поиском похожих конференций из-за более представительного опорного множества авторов.

ЗАКЛЮЧЕНИЕ

Разработанные авторами алгоритмы тематического поиска конференций позволяют на основе наукометрических данных производить подбор интересующих конечного пользователя конференций с учетом их авторитетности в научном сообществе. Разработанные алгоритмы поиска по образцу нечувствительны к

языку и не требуют наличия полнотекстовых описаний докладов и материалов конференций.

СПИСОК ЛИТЕРАТУРЫ

1. Козицын А.С., Афонин С.А. Метод поиска экспертов по данным наукометрических систем // Электронные библиотеки. 2021. Т. 24. № 5. С. 870–888.
 2. Зельдина М.М. Рецензирование. М.: ООО «Ваше цифровое издательство», 2019. 30 с.
URL:https://elpub.ru/images/knowledge-base/metod_neicon/Рецензирование.pdf
 3. Козицын А.С., Афонин С.А., Шачнев Д.А. Использование методов тематического анализа в наукометрических системах // Электронные библиотеки. 2021. Т. 24. № 2. С. 315–338.
 4. Vasenin V., Itkes A., Krivchikov M., Yavtushenko E. Chrelbac data access control model for large-scale interactive informational-analytical systems // Journal of Computer Virology and Hacking technique. 2020. P. 1–19.
 5. Садовничий В.А., Васенин В.А., Афонин С.А., Козицын А.С., Голомазов Д.Д., Информационная система Истина как big data – инструментарий в области управления на основе анализа наукометрических данных // Знания – Онтологии – Теории (ЗОНТ-2015). Материалы Всероссийской конференции с международным участием. Российская академия наук, Сибирское отделение; Институт математики им. С.Л. Соболева. 2015. С. 115–123.
 6. Тарасевич Ю.Ю., Шиняева Т.С. Наука самолетопоклонников: 40 лет спустя // Cloud of Science. 2017. Т. 4. № 4. С. 525–547.
 7. Соколов А.В. Что есть информационная потребность? // Труды Санкт-Петербургского государственного университета культуры и искусств. 2013. Т. 197. С. 7–18.
 8. Oracle Text In Oracle Database 12 с.
URL:<https://www.oracle.com/technetwork/database/12coracletexttwp-1961244.pdf>
 9. Введение в полнотекстовый поиск в PostgreSQL.
URL: http://www.sai.msu.su/~megera/postgres/talks/fts_pgsql_intro.htm
 10. Менщиков А. Эффективный поиск на сайте с помощью Elasticsearch. URL: <https://codex.so/elastic-search>
 11. Родичев Е. Приручаем Elasticsearch.
-

URL: <https://club.directum.ru/post/360100>

12. Козицын А.С., Афонин С.А. Алгоритм разрешения неоднозначности имен авторов в ИАС ИСТИНА // Современные информационные технологии и ИТ-образование. 2020. Т. 16. № 1. С. 108–117.

13. Kozitsin A., Afonin S., Shachnev D. Algorithm for searching experts in scientometric systems // CEUR Workshop Proceedings. Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). 2022. Vol. 3066. P. 59–68.

14. Shachnev D.A. Searching for activity results and experts in a given subject area, taking results significance into account // Программная инженерия. 2021. Т. 12. №5. С. 260–266.

15. Козицын А.С., Афонин С.А. Разрешение неоднозначностей при определении авторов публикации с использованием графов соавторства в больших коллекциях библиографических данных // Программная инженерия. 2017. Т. 8. № 12. С. 556–562.

16. Козицын А.С., Афонин С.А., Шачнев Д.А. Алгоритм поиска по ключевым словам специалистов в заданной предметной области // Современные информационные технологии и ИТ-образование. 2021. Т. 17. № 1. С. 124–133.

17. Васенин В.А., Занчурич М.А., Козицын А.С., Кривчиков М.А., Шачнев Д.А. Архитектурно-технологические аспекты разработки и сопровождения больших информационно-аналитических систем в сфере науки и образования // Программная инженерия. 2017. Т. 8. №10. С. 448–455.

18. Васенин В.А., Афонин С.А., Козицын А.С., Голомазов Д.Д., Бахтин А.В., Ганкин Г.М. Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА) // Обзорение прикладной и промышленной математики. 2012. Т. 19. № 2. С. 239–240.

19. Vasenin V.A., Zanchurin M.A., Kozitsyn A.S., Krivchikov M.A., Shachnev D.A. Architectural and technological aspects of the cloud data analysis system development, case of istina system // CEUR Workshop Proceedings. 2017. "APSSE 2017 – Proceedings of the 5th International Conference on Actual Problems of System and Software Engineering". P. 90–96.

20. Afonin S., Kozitsyn A., Astapov I. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529–534.

21. Козицын А.С., Афонин С.А., Шачнев Д.А. Метод оценки тематической близости научных журналов // Программная инженерия. 2020. № 6. С. 335–341.

USING THE CO-AUTHORITY GRAPH FOR THE THEMATIC SEARCH OF CONFERENCES ON SCIENTOMETRIC DATA

A.S. Kozitsyn¹ [0000-0002-8065-9061], S.A. Afonin² [0000-0003-3058-9269],
D.A. Shachnev³ [0000-0002-5940-9180]

Institute of Mechanics Lomonosov Moscow State University

¹*alexanderkz@mail.ru*, ²*serg@msu.ru*, ³*mitya57@gmail.com*

Abstract

Thematic information search is used in various fields of activity. The use of thematic analysis tools to search for conferences allows you to increase the completeness of the search and coverage of conferences, helps to expand the circle of scientific communication of young scientists and the formation of closer scientific connections. The search algorithms developed by the authors use the co-authorship graph and the reference set of authors. The set can be obtained using methods of thematic search of experts or based on given samples. The developed algorithms are language insensitive and take into account the authority of conferences in the scientific community. Approximation was carried out on the data of the scientometric system IAS ISTINA.

Keywords: *thematic search, bibliographic data, conference search, co-authorship graph, information systems, scientometrics*

REFERENCES

1. *Kozitsyn A.S., Afonin S.A.* Metod poiska ekspertov po dannym nauko-metricheskikh sistem // Elektronnye biblioteki. 2021. T. 24. № 5. S. 870–888.
2. *Zeldina M.M.* Retsenzirovaniye. M.: OOO «Vashe tsifrovoye izdatelstvo», 2019. S. 30.
URL:https://elpub.ru/images/knowledge-base/metod_neicon/Retsenzirovaniye.pdf
3. *Kozitsyn A.S., Afonin S.A., Shachnev D.A.* Ispolzovaniye metodov tematicheskogo analiza v nauko-metricheskikh sistemakh // Elektronnye biblioteki. 2021. T. 24. № 2. S. 315–338.
4. *Vasenin V., Itkes A., Krivchikov M., Yavtushenko E.* Chrelybac data access control model for large-scale interactive informational-analytical systems // Journal of Computer Virology and Hacking technique. 2020. P. 1–19.
5. *Sadovnichii V.A., Vasenin V.A., Afonin S.A., Kozitsyn A.S., Golomazov D.D.* Informatsionnaya sistema "ISTINA" kak big data – instrumentarii v oblasti upravleniya na osnove analiza nauko-metricheskikh dannykh // Znaniya–Ontologii–Teorii (ZONT-2015). Materialy Vserossiyskoy konferentsii s mezhdunarodnym uchastiem. Rossiyskaya akademiya nauk, Sibirskoye otdeleniye; Institut matematiki im. S.L. Soboleva. 2015. S. 115–123.
6. *Tarasevich Iu.Iu., Shiniayeva T.S.* Nauka samoletopoklonnikov: 40 let spustia // Cloud of Science. 2017. T. 4. № 4. C. 525–547.
7. *Sokolov A.V.* Chto est informatsionnaya potrebnost? // Trudy Sankt-Peterburgskogo gosudarstvennogo universiteta kultury i iskusstv. 2013. T. 197. S. 7–18.
8. Oracle Text In Oracle Database 12 c.
URL:<https://www.oracle.com/technetwork/database/12coracletexttp-1961244.pdf>
9. Vvedeniye v polnotekstovyy poisk v PostgreSQL.
URL: http://www.sai.msu.su/~megera/postgres/talks/fts_pgsql_intro.htm
10. *Menshchikov A.* Effektivnyy poisk na saite s pomoshchiu Elasticsearch.
URL:<https://codex.so/elastic-search>
11. *Rodichev E.* Priruchayem Elasticsearch.
URL: <https://club.directum.ru/post/360100>
12. *Kozitsyn A.S., Afonin S.A.* Algoritm razresheniya neodnoznachnosti imen avtorov v IAS ISTINA // Sovremennyye informatsionnyye tekhnologii i IT-obrazovaniye. 2020. T. 16. № 1. S. 108–117.

13. *Kozitsin A., Afonin S., Shachnev D.* Algorithm for searching experts in scientometric systems // CEUR Workshop Proceedings. Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). 2022. Vol. 3066. P. 59–68.

14. *Shachnev D.A.* Searching for activity results and experts in a given subject area, taking results significance into account // *Programmnaia inzheneriia*. 2021. T. 12, № 5. S. 260–266.

15. *Kozitsyn A.S., Afonin S.A.* Razreshenie neodnoznachnostei pri opredelenii avtorov publikatsii s ispolzovaniem grafov soavtorstva v bolshikh kolleksiakh bibliograficheskikh dannykh // *Programmnaia inzheneriia*. 2017. T. 8. № 12. S. 556–562.

16. *Kozitsyn A.S., Afonin S.A., Shachnev D.A.* Algoritm poiska po kliuchevym slovam spetsialistov v zadannoi predmetnoi oblasti // *Sovremennye informatsionnye tekhnologii i IT-obrazovanie*. 2021. T. 17. № 1. S. 124–133.

17. *Vasenin V.A., Zanchurin M.A., Kozitsyn A.S., Krivchikov M.A., Shachnev D.A.* Arkhitekturno-tekhnologicheskie aspekty razrabotki i soprovozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obrazovaniia // *Programmnaia inzheneriia*. 2017. Tom 8, № 10. S. 448–455.

18. *Vasenin V.A., Afonin S.A., Kozitsyn A.S., Golomazov D.D., Bakhtin A.V., Gan'kin G.M.* Intellektualnaia sistema tematicheskogo issledovaniia nauchno-tekhnicheskoi informatsii (ISTINA) // *Obozrenie prikladnoi i promyshlennoi matematiki*. 2012. T. 19. № 2. S. 239–240

19. *Vasenin V.A., Zanchurin M.A., Kozitsyn A.S., Krivchikov M.A., Shachnev D.A.* Architectural and technological aspects of the cloud data analysis system development, case of istina system // CEUR Workshop Proceedings. "APSSE 2017 – Proceedings of the 5th International Conference on Actual Problems of System and Software Engineering". 2017. S. 90–96.

20. *Afonin S., Kozitsyn A., Astapov I.* Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529–534.

21. *Kozitsyn A.S., Afonin S.A., Shachnev D.A.* Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov // *Programmnaia inzheneriia*. 2020. № 6. S. 335–341.

СВЕДЕНИЯ ОБ АВТОРАХ



КОЗИЦЫН Александр Сергеевич – ведущий научный сотрудник, к. ф.-м. н., окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области информационного поиска и баз данных.

Alexander Sergeevich KOZITSYN – Leading Researcher, Ph.D., graduated from M.V. Lomonosov Moscow State University. Specialist in the field of information retrieval and database.

email: alexanderkz@mail.ru,

ORCID: 0000-0002-8065-9061



АФОНИН Сергей Александрович – ведущий научный сотрудник, к. ф.-м. н., окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области регулярных языков и информационных систем.

Sergey Alexandrovich AFONIN – Leading Researcher, Ph.D., graduated from M.V. Lomonosov Moscow State University. Specialist in the field of regular languages and information systems.

email: serg@msu.ru,

ORCID:0000-0003-3058-9269



Шачнев Дмитрий Алексеевич – программист, окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области информационных систем.

Dmitry Alekseevich SHACHNEV – programmer, graduated from M.V. Lomonosov Moscow State University. Specialist in information systems.

email: mitya57@gmail.com,

ORCID: 0000-0002-5940-9180

Материал поступил в редакцию 22 декабря 2022 года

УДК 004.021; 004.42

СЕМАНТИЧЕСКОЕ АННОТИРОВАНИЕ МАТЕМАТИЧЕСКИХ ФОРМУЛ В PDF-ДОКУМЕНТАХ

О. А. Невзорова¹ [0000-0001-8116-9446], К. С. Николаев² [0000-0003-3204-238X]

^{1,2}Казанский (Приволжский) федеральный университет, ул. Кремлевская, 35,
г. Казань, 420008

¹onevzoro@gmail.com, ²konnikolaeff@yandex.ru

Аннотация

Дан обзор существующих решений по семантическому анализу математических документов, а также описан метод автоматического семантического анализа документов, представленных в формате PDF. Разработанный метод позволяет выделять математические формулы внутри документа, анализировать их структуру, выполнять поиск локальных переменных формулы и их определений в документе, а также связывать переменные формулы и понятия из онтологии. Преимуществом разработанного метода перед другими существующими является независимость от разметки исходного PDF-документа, что расширяет область применения метода. Приведены оценки полноты, точности и F-меры для алгоритмов поиска переменных и связывания локальных переменных с формулами. Полученная семантическая разметка документа позволяет создавать коллекции документов, пригодных для сервиса семантического поиска формул, который является одним из сервисов цифровой библиотеки Lobachevskii-DML.

Ключевые слова: семантический анализ, PDF, обработка документов, научные журналы, Lobachevskii-DML

ВВЕДЕНИЕ

Семантический поиск ориентирован на поиск в цифровых коллекциях таких публикаций, которые представляют собой документы с семантической разметкой компонентов текста. Математические тексты отличаются высокой структурированностью, наличием компонентов фиксированной семантики, таких как теоремы, доказательства, формулы.

Задача поиска документов по математическим формулам является актуальной для проведения научных исследований, подготовки статей, изучения математических дисциплин. В работе [1] описан семантический поисковик по математическим формулам, который использует набор данных, построенный по коллекции научных статей журнала «Известия ВУЗов. Математика» за 1997–2009 гг. В настоящей статье представлены новые улучшенные алгоритмы для построения набора данных для семантического поиска по математическим формулам, что позволит качественно улучшить результаты поиска.

Математический поиск по формулам можно разделить на две категории – поиск формул по структуре и по содержанию. Поиск формул по структуре сводится к получению списка формул, которые частично или полностью совпадают со структурой формулы, заданной в поисковом запросе. Такой подход не учитывает семантики формул.

Более эффективным, но сложным в реализации является поиск математических статей по содержанию формулы. Для определения содержания формулы необходимо выделить переменные, входящие в ее состав, и определить математические понятия, обозначаемые переменными. Дополнительные сложности вносит разнообразие шаблонов оформления математических документов для разных информационных систем научных журналов. В настоящее время наиболее популярными форматами представления математических формул в научных статьях являются: графическое изображение (статьи в формате pdf); формулы в редакторе Microsoft Word; формат LaTeX; формат MathML. Коллекции статей в цифровых математических библиотеках представлены преимущественно в формате PDF. Распознавание текста и, в частности, математических выражений является основной задачей данного исследования. Математические формулы, извлекаемые при распознавании в научных статьях, и их описания являются исходными данными для построения улучшенного набора данных для математического поисковика.

Статья организована следующим образом. Вначале приведен обзор существующих решений, предназначенных для анализа структуры и семантики математических документов с формулами. Затем представлены новые алгоритмы семантического аннотирования математических формул, извлеченных из статей в

формате pdf. В заключении приведены оценки построенных алгоритмов и общие выводы.

СУЩЕСТВУЮЩИЕ РЕШЕНИЯ ДЛЯ АНАЛИЗА СТРУКТУРЫ И СЕМАНТИКИ МАТЕМАТИЧЕСКИХ ДОКУМЕНТОВ

Существующие методы анализа структуры и семантики математических документов можно разделить на две категории. Методы первой группы ориентированы на извлечение метаданных, анализ логической структуры документа и извлечение внутритекстовых ссылок и цитирований. Методы второй группы дополнительно анализируют определения и формулы, присутствующие в тексте документа. Приведем некоторые работы из первой группы.

В работе [2] авторы предлагают протокол семантической разметки научных документов. Этот протокол основан на обработке XML, сегментации и семантической разметке текста. Наибольшее внимание авторы уделяют библиографии, внутритекстовым ссылкам и названиям разделов. Результатом работы является набор связанных данных для проекта Linked Open Data¹.

Авторы работы [3] также предлагают метод семантической аннотации научных документов. Для обработки документов используется программа PDFX, преобразующая PDF-документ в формат XML. Основными извлекаемыми данными в этой работе является расширенная информация о внутритекстовых цитатах.

В статье [4] авторы описывают процесс семантической разметки сборников CEUR-WS. Для каждого документа формируется информация о логической структуре документа, семантике текста (с помощью классификаторов SVM) и выполняется преобразование в формат RDF для последующей публикации в DBPedia².

В [5] приведен метод извлечения информации из документов в форматах PDF и XML. Из статей извлекается информация о логической структуре документа и вспомогательных материалах документа (базовые метаданные, информация о финансировании, названия таблиц и изображений, названия проекта и др.).

¹ <https://lod-cloud.net/>

² <https://www.dbpedia.org/>

Работы второй группы используют различные подходы (основанные на правилах, векторные представления текста, машинное обучение) для поиска определений математических формул в тексте. К примеру, в [6] авторами использована модель векторного представления слов для извлечения структуры математических документов и определений формул. Такой подход неплохо работает для извлечения структурных компонентов документов, но показывает нелучшие результаты при извлечении определений формул. Авторы указывают, что, по их данным, лишь для 70% всех формул в математических документах явно даются их определения [7].

В [8] обсуждается подход к извлечению формул из документов в формате LaTeX и представлению формул в специальном формате для математического поиска. Кроме того, данный метод способен анализировать текстовые документы с формулами в формате MathML. При этом связывание формул с их определениями не проводится.

В работе [9] проведены эксперименты по учету текстового контекста математических формул для повышения качества их преобразования в машиночитаемые форматы. Важным результатом этой работы является датасет, в котором размечены формулы, их компоненты и понятия, встречающиеся в тексте.

В [10] рассмотрен подход к аннотированию формул в формате XML на основе набора лингвистических шаблонов для установления легенды формульной переменной в тексте.

Большинство методов из работ, упомянутых выше, сильно зависит от входного документа, его формата и структуры. В настоящей статье предложен универсальный метод определения структуры документа в формате PDF и связывания переменных в тексте с главными формулами для определения семантического содержимого документа. Набор данных, построенный на основе разработанных алгоритмов, будет использован для улучшения качества семантического поиска математических формул.

МЕТОД СЕМАНТИЧЕСКОГО АННОТИРОВАНИЯ ФОРМУЛ В PDF-ДОКУМЕНТЕ

Семантическое аннотирование формулы заключается в выделении из текста математической статьи формулы, отвечающей специальным требованиям, с последующим анализом ее структурных элементов и связыванием выделенных переменных формулы с легендами, данными в текстовом контексте формулы. Схема разработанного алгоритма семантического аннотирования приведена на Рис. 1.

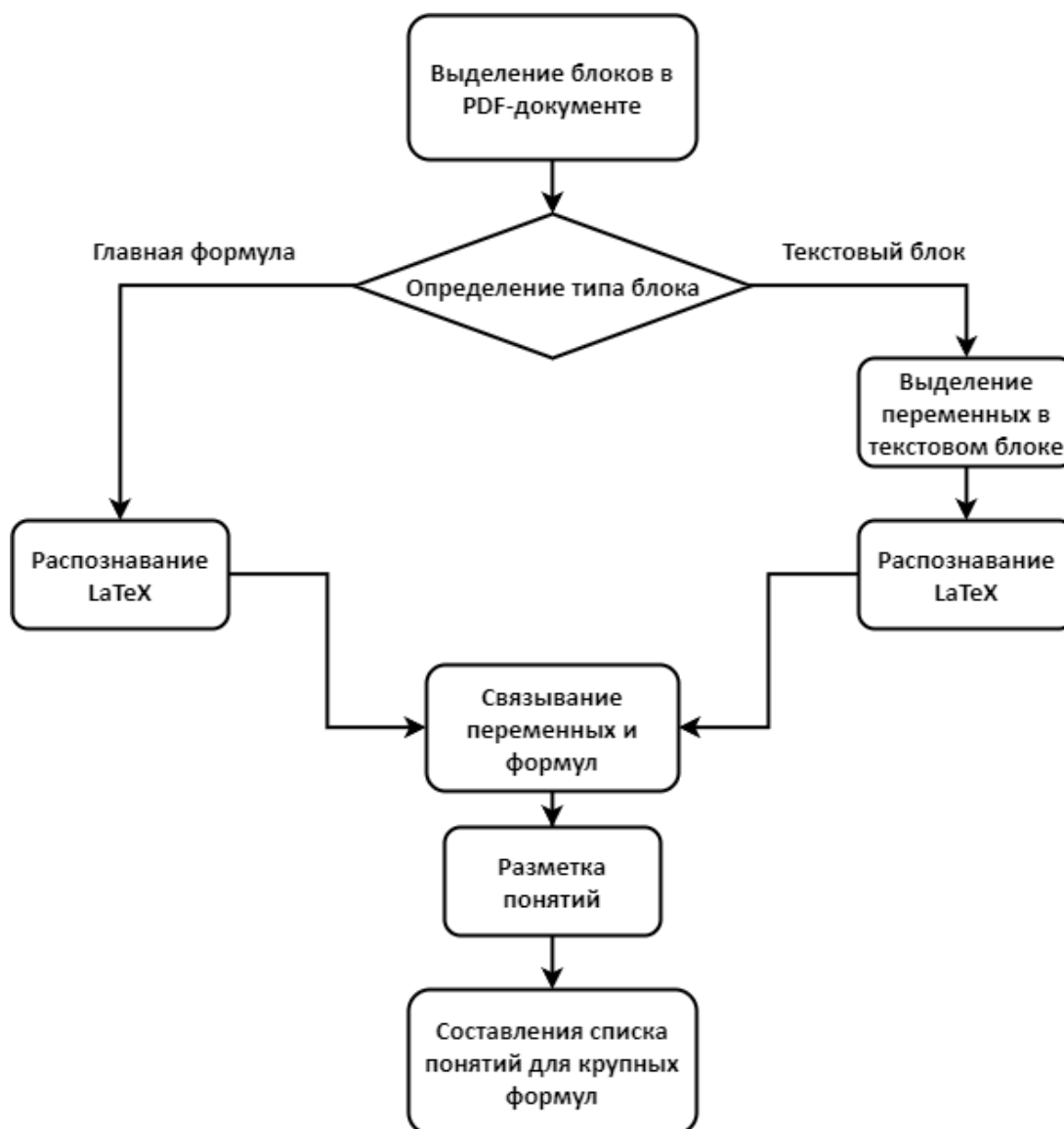


Рис. 1. Схема алгоритма семантического аннотирования PDF-документов

Главная задача семантического аннотирования формул заключается в разработке программного решения, позволяющего выделить набор переменных в формулах математического документа и связать переменные с математическими понятиями, используя математическую онтологию. Полученная семантическая разметка документа позволит создать коллекцию документов, пригодных для сервиса семантического поиска формул, являющегося частью набора сервисов цифровой платформы Lobachevskii-DML.

В качестве тестовой коллекции мы используем набор случайно выбранных документов, размещенных на портале MathNet.ru. Документы в коллекции представлены в виде текстового PDF, что заметно упрощает анализ документа. Примеры, использованные в настоящей работе, содержатся в статьях Р. Р. Кучарова, Ю. Х. Эшкабилова, Ю. Г. Никонорова, А. И. Парфенова³. Для решения задачи семантического аннотирования формул в PDF-документах необходимо решить следующие задачи:

1. Разделение документа на блоки.
2. Выделение главных формул и текстовых блоков.
3. Поиск переменных в текстовых блоках.
4. Распознавание главных формул и локальных переменных.
5. Связывание формул и локальных переменных.
6. Разметка математических понятий в текстовых блоках на основе онтологии OntoMathPro⁴.
7. Связывание выделенных понятий с переменными формулы.

Опишем подробнее каждую из указанных подзадач.

Разделение документа на блоки. Под блоками будем понимать обособленные участки символов, такие как формулы и текстовые абзацы. Данный этап выполняется с помощью функционала анализа разметки документа, встроенного в библиотеку `pdfminer`, и группировки отдельных элементов в блоки. На Рис. 5 приведен пример результата разделения страницы документа на блоки. При этом

³ <http://mi.mathnet.ru/mt270>, <http://mi.mathnet.ru/mt271>, <http://mi.mathnet.ru/mt272>

⁴ <https://github.com/CLLKazan/OntoMathPro>

для последующих этапов обработки документа сохраняется информация об отдельных строках документа и символах в строках. Кроме того, выполняется фильтрация служебных блоков, таких как номер формулы (например, «(1)»), и номеров страниц.

Введем понятия главной формулы, текстового блока и локальной переменной текстового блока.

Под главной формулой будем понимать формулу, отделенную от текстовой информации разрывами строк (например, такую, как на Рис. 2).

Текстовым блоком называется блок, содержащий как текстовую, так и формульную информацию. Примеры текстовых блоков приведены на Рис. 3.

Под локальной переменной текстового блока понимаются формулы, расположенные внутри текстовых блоков. Примеры локальных переменных приведены на Рис. 4

Выделение главных формул и текстовых блоков. Для дальнейшего анализа структуры документа необходимо различать блоки с текстом и главные формулы (такие, как формула $H = H_0 - (T_1 + T_2)$ на Рис. 5). Для этого применяется проверка количества слов в блоке с помощью регулярного выражения. В случае, если количество слов в блоке меньше порогового, блок считается главной формулой. На этом этапе алгоритма для главных формул сохранялась соответствующая часть изображения страницы из исходного документа. Например, для формулы $H = H_0 - (T_1 + T_2)$ на Рис. 5 сохраняется изображение, приведенное на Рис. 6.

Пусть $u(x)$ и $v(y)$ — неотрицательные непрерывные функции на Ω_1 и Ω_2 , $0 \in \text{Ran}(u) \cap \text{Ran}(v)$, $k_0(x, y) = u(x)v(y)$ и $\varphi_j(\cdot)$ — вещественнозначные непрерывные функции на Ω_j , для которых выполняются следующие равенства:

$$\int_{\Omega_j} \varphi_j(\xi) d\mu_j(\xi) = 0, \quad \int_{\Omega_j} \varphi_j^2(\xi) d\mu_j(\xi) = 1, \quad j = 1, 2.$$

В этом параграфе мы рассмотрим вопросы о существовании отрицательных собственных значений, лежащих ниже нижнего края ЧИО H вида (1), и их количестве при вышеуказанных предположениях.

Рис. 2. Пример главной формулы в документе (выделена цветным

прямоугольником)

Здесь и в дальнейшем будем предполагать, что $ds = d\mu_1(s)$, $dt = d\mu_2(t)$ и $\mu_1(\Omega_1) = \mu_2(\Omega_2) = 1$.

Пусть $k_0(x, y)$ – произвольная вещественнозначная непрерывная функция на $\Omega_1 \times \Omega_2$. Обозначим оператор умножения на функцию $k_0(x, y)$ через V_0 , т. е.

$$(V_0 f)(x, y) = k_0(x, y) f(x, y).$$

Рассмотрим линейный ограниченный самосопряженный ЧИО

$$V = V_0 - A, \tag{2}$$

действующий в пространстве $L_2(\Omega_1 \times \Omega_2)$, где $A = A_1 + A_2$.

Рис. 3. Примеры текстового блока (выделен цветным прямоугольником)

Здесь и в дальнейшем будем предполагать, что $ds = d\mu_1(s)$ $dt = d\mu_2(t)$ и $\mu_1(\Omega_1) = \mu_2(\Omega_2) = 1$.

Пусть $k_0(x, y)$ – произвольная вещественнозначная непрерывная функция на $\Omega_1 \times \Omega_2$. Обозначим оператор умножения на функцию $k_0(x, y)$ через V_0 т. е.

$$(V_0 f)(x, y) = k_0(x, y) f(x, y).$$

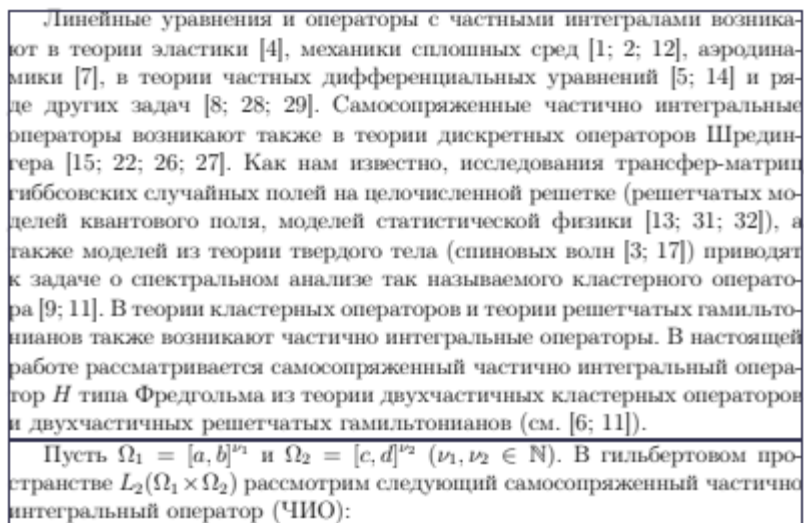
Рассмотрим линейный ограниченный самосопряженный ЧИО

$$V = V_0 - A, \tag{2}$$

действующий в пространстве $L_2(\Omega_1 \times \Omega_2)$ где $A = A_1 + A_2$

Рис. 4. Примеры локальных переменных (выделены цветным прямоугольником)

На Рис. 7 приведен результат определения типа блока – синим цветом выделены текстовые блоки, красным – блоки с главной формулой.



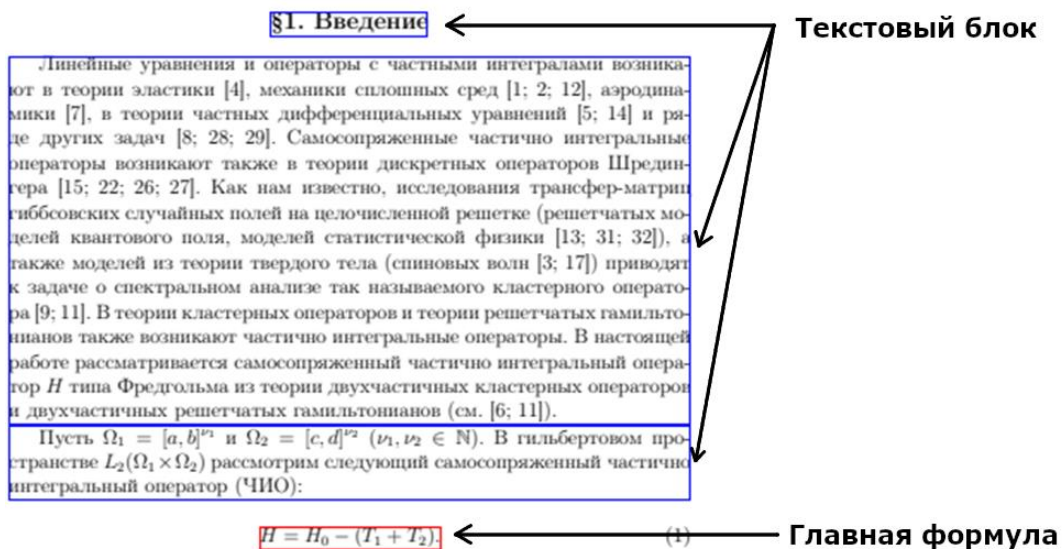
$$H = H_0 - (T_1 + T_2). \tag{1}$$

© Р. Р. Кучаров, Ю. Х. Эшкабилов; 2014

Рис. 5. Пример разделения страницы документа на блоки

$$H = H_0 - (T_1 + T_2)$$

Рис. 6. Пример изображения, привязанного к блоку с главной формулой



© Р. Р. Кучаров, Ю. Х. Эшкабилов; 2014

Рис. 7. Определение типа блока

Поиск переменных в текстовых блоках. Для выделения переменных в текстовых блоках выполняются сегментация текста на предложения и затем токенизация (выделение слов) в предложениях. В список локальных переменных добавляются все участки текста, не входящие в список слов текстового блока. Схема алгоритма поиска приведена на Рис. 8.

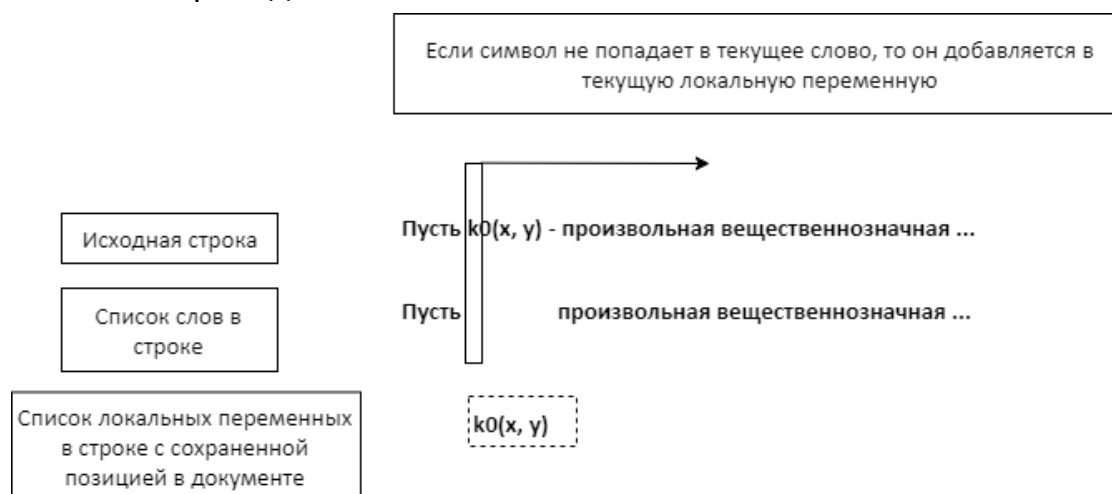


Рис. 8. Поиск локальных переменных в текстовых блоках

В результате этой операции формируется набор локальных переменных, и для каждой собирается набор символов, входящих в имя переменной. Однако некоторые специфические математические символы (например, символ интеграла, символ суммы) не распознаются библиотекой pdfminer. По этой причине для каждой локальной переменной дополнительно извлекается соответствующая часть изображения страницы из исходного документа. На Рис. 9 прямоугольниками зеленого цвета отмечены части изображения страницы, соответствующие локальным переменным.

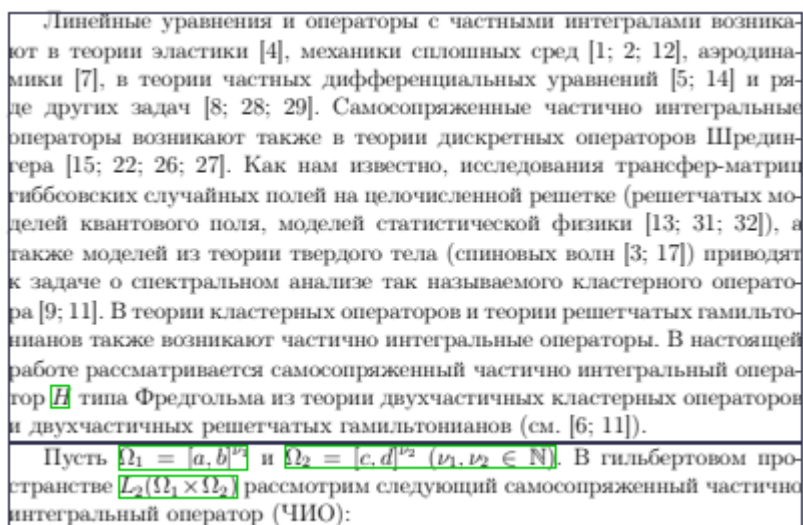


Рис. 9. Локальные переменные и их позиции на изображении страницы (отмечены прямоугольниками зеленого цвета)

Распознавание главных формул и локальных переменных. Для более корректного связывания переменных в тексте с главными формулами выполняется преобразование главных формул и локальных переменных в формат LaTeX. Для этого применяется библиотека `pix2tex`, использующая предобученную нейросетевую модель для распознавания формул LaTeX на изображениях. В Таблице 1 приведены результаты распознавания с указанием исходного текста формулы и переменных в документе и распознанного LaTeX представления. Распознавание не всегда происходит корректно, но для успешного связывания переменной и формулы нет необходимости в идеальном совпадении переменной и части формулы.

Связывание формул и локальных переменных производилось с помощью выделения обозначений формул с учетом нижних и верхних индексов (например, $f(x)$, $f_0(x, y)$) и поиска совпадающих формул в главной формуле и локальных переменных, находящихся в пределах некоторого текстового окна до и после формулы. Учет расстояния между переменной и формулой был введен по причине того, что некоторые переменные могут переопределяться автором в различных частях документа. Для переменных, не обозначающих формулу с аргументами (например, на Рис. 9 выделен частично интегральный оператор H), проводится прямой поиск в главной формуле. На Рис. 10 Рис. 10 приведена схема алгоритма связывания главных формул с локальными переменными. На Рис. 11 дан пример

связывания главной формулы с локальными переменными.

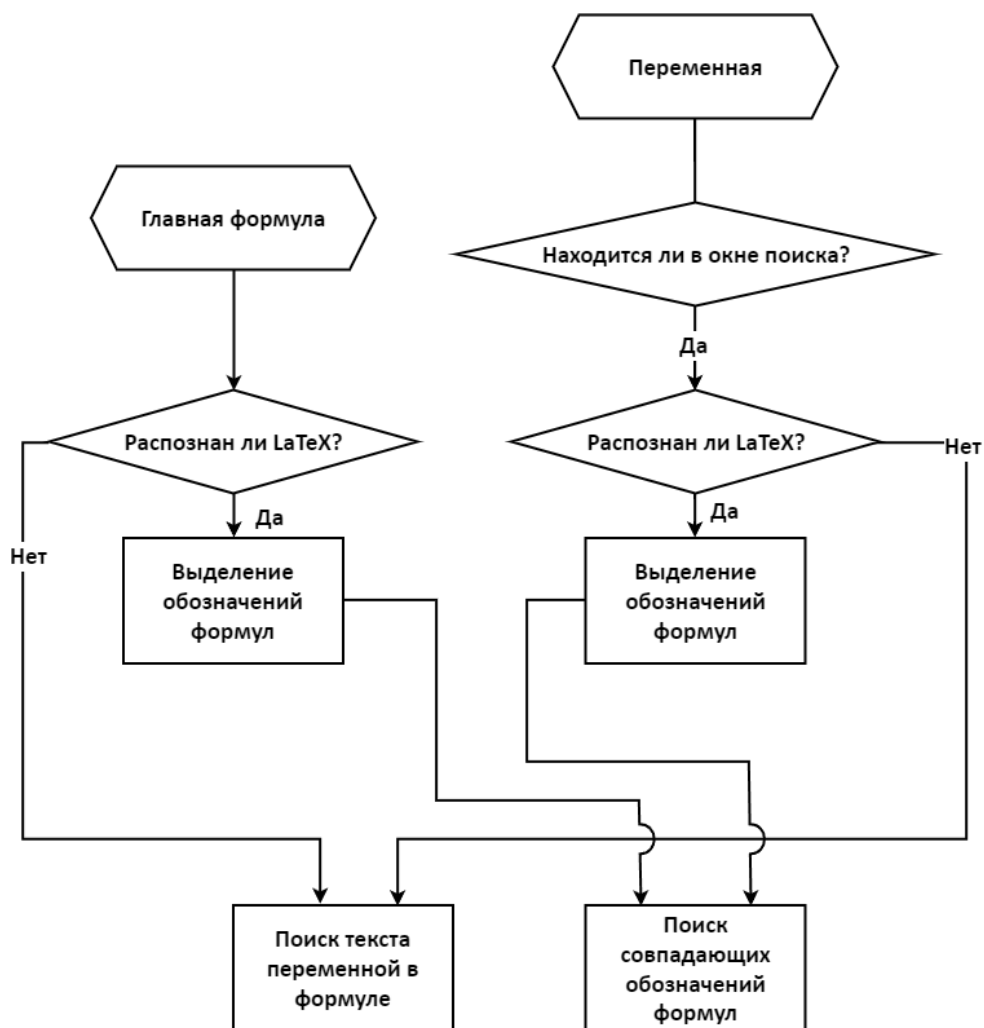


Рис. 10. Схема алгоритма связывания локальных переменных и главных формул

§1. Введение

Линейные уравнения и операторы с частными интегралами возникают в теории эластики [4], механики сплошных сред [1; 2; 12], аэродинамики [7], в теории частных дифференциальных уравнений [5; 14] и ряде других задач [8; 28; 29]. Самосопряженные частично интегральные операторы возникают также в теории дискретных операторов Шредингера [15; 22; 26; 27]. Как нам известно, исследования трансфер-матриц гиббсовских случайных полей на целочисленной решетке (решетчатых моделей квантового поля, моделей статистической физики [13; 31; 32]), а также моделей из теории твердого тела (спиновых волн [3; 17]) приводят к задаче о спектральном анализе так называемого кластерного оператора [9; 11]. В теории кластерных операторов и теории решетчатых гамильтонианов также возникают частично интегральные операторы. В настоящей работе рассматривается самосопряженный частично интегральный оператор L -типа Фредгольма из теории двухчастичных кластерных операторов и двухчастичных решетчатых гамильтонианов (см. [6; 11]).

Пусть $\Omega_1 = [a, b]^{\nu_1}$ и $\Omega_2 = [c, d]^{\nu_2}$ ($\nu_1, \nu_2 \in \mathbb{N}$). В гильбертовом пространстве $L_2(\Omega_1 \times \Omega_2)$ рассмотрим следующий самосопряженный частично интегральный оператор (ЧИО):

$$H = H_0 - (T_1 + T_2), \tag{1}$$

Рис. 11. Связывание главной формулы и переменной

Таблица 1. Примеры распознанных формул и переменных

Исходное текстовое представление формулы	Распознанное представление в LaTeX
$H = H_0 - (T_1 + T_2)$	$H=H_{\{0\}}-(T_{\{1\}}+T_{\{2\}})$
$H_0 f(x, y) = k_0(x, y) f(x, y)$	$H_{\{0\}}f(x,y)=k_{\{0\}}(x,y)f(x,y)$
$\Omega_1 \times \Omega_2$	$\backslash\Omega_{\{1\}}\backslash\times\backslash\Omega_{\{2\}}$
$k_2(x, t, y) f(x, t) dt$	$k_{\{2\}}(x,t,y)f(x,t)d u$

В Таблице 2 приведены оценки полноты и точности алгоритма связывания локальных переменных и главных формул. В качестве экспериментальной выборки были использованы 9 документов на портале Math-Net.ru. Показатель полноты алгоритма связывания указывает на необходимость дальнейшего улучшения процедуры связывания локальных переменных и главных формул.

Таблица 2. Оценки полноты и точности для алгоритма связывания локальной переменной и главной формулы

Корректные связывания (TP)	Некорректные связывания (FP)	Пропущенные связывания (FN)	Точность	Полнота	F-мера
38	6	14	0,86	0,73	0,79
30	4	49	0,88	0,38	0,53
51	24	24	0,68	0,68	0,68
48	6	42	0,89	0,53	0,67
110	60	35	0,65	0,76	0,70
68	6	30	0,92	0,69	0,79
98	12	6	0,89	0,94	0,92
100	32	6	0,76	0,94	0,84
90	24	18	0,79	0,83	0,81
Средние значения			0,81	0,72	0,75

Разметка математических понятий в текстовых блоках на основе онтологии OntoMathPro

Для определения понятия, связанного с локальной формулой, был доработан и применен метод автоматической разметки текстовых документов понятиями из математической онтологии. Этот метод применяется при подготовке образовательных курсов по математике в Казанском федеральном университете. Основная идея данного алгоритма заключается в поиске всех цепочек слов в предложении и сравнении их с аналогично построенными цепочками в понятиях онтологии. Алгоритм принимает на вход документы в формате html, поэтому для его применения в данной задаче был создан метод генерации промежуточного html-представления PDF-документа. В тегах такого представления было получено текстовое представление исходного документа с указанием порядкового номера блока. Подробная схема алгоритма разметки математических документов приведена на Рис. 12.

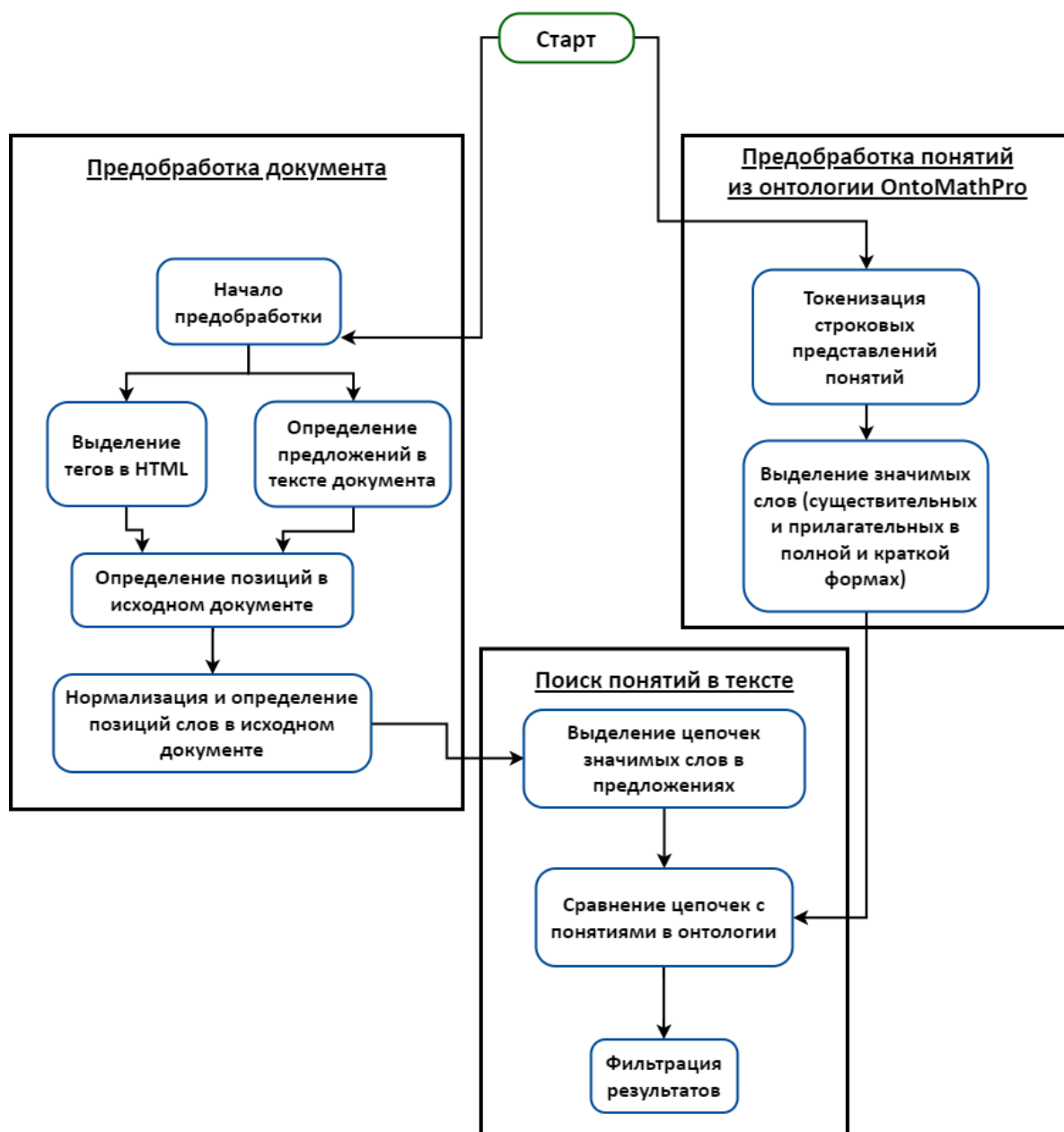


Рис. 12. Схема алгоритма разметки математических документов

Для применения метода разметки математических документов в выделении понятий в текстовых блоках PDF-документов была применена онтология профессиональной математики OntoMathPro [11].

Онтология профессиональной математики OntoMathPro – это прикладная онтология для автоматической обработки профессиональных математических статей на русском и английском языках, разработанная в Казанском федеральном университете. Эта онтология охватывает широкий спектр областей математики,

таких как теория чисел, теория множеств, алгебра, анализ, геометрия, теория вычислений, дифференциальные уравнения, численный анализ, теория вероятностей и статистика. Каждый концепт онтологии имеет аннотацию, имя на русском и английском языках, включая синонимы. В онтологии можно выделить две таксономии по отношению *ISA* – иерархия областей математики и иерархия объектов математического знания (Рис. 13). Понятия онтологии, включенные в иерархию объектов математического знания, используются в разработанном алгоритме связывания переменных формулы.



Рис. 13. Фрагмент онтологии OntoMathPro: разделы математики и элементы математического знания

Связывание выделенных понятий с переменными формулы

В результате работы метода разметки математических документов формируется список понятий и позиций слов в документе, к которым привязано соответствующее понятие из онтологии OntoMathPro. Кандидаты на связывание с переменной выбираются в пределах предложения, в котором находится переменная. На данном этапе к переменной привязываются непосредственно соседние понятия в тексте. На Рис. 14 приведен пример разметки текста понятиями из онтологии OntoMathPro. В скобках после каждого цветного выделения приведено название распознанного понятия.

Линейные уравнения (линейное уравнение) и операторы(оператор) с частными интегралами(интеграл) возникают в теории эластики [4], механики сплошных сред [1; 2; 12], аэродинамики [7], в теории частных дифференциальных уравнений(дифференциальное уравнение) [5; 14] и ряде других задач [8; 28; 29]. Самосопряженные (сопряженный оператор) частично интегральные операторы (интегральный оператор) возникают также в теории дискретных операторов Шредингера [15; 22; 26; 27]. Как нам известно, исследования трансфер-матриц гиббсовских случайных полей (поле случайное) на целочисленной решетке (решетка) (решетчатых моделей квантового поля (поле), моделей статистической (статистическая модель) физики [13; 31; 32]), а также моделей из теории твердого тела (тело) (спиновых волн [3; 17]) приводят к задаче о спектральном анализе так называемого кластерного оператора [9; 11]. В теории кластерных операторов (оператор) и теории решетчатых гамильтонианов (гамильтониан) также возникают частично интегральные операторы. В настоящей работе рассматривается самосопряженный (сопряженный оператор) частично интегральный оператор (интегральный оператор) H типа Фредгольма из теории двухчастичных кластерных операторов (оператор) и двухчастичных решетчатых гамильтонианов (гамильтониан) (см. [6; 11]).

$$H = H_0 - (T_1 + T_2). \quad (1)$$

Пусть $\Omega_1 = [a, b]^{\nu_1}$ и $\Omega_2 = [c, d]^{\nu_2}$ ($\nu_1, \nu_2 \in \mathbb{N}$). В гильбертовом пространстве (гильбертово пространство) $L_2(\Omega_1 \times \Omega_2)$ рассмотрим следующий самосопряженный (самосопряженный оператор) частично интегральный оператор (интегральный оператор) (ЧИО):

Рис. 14. Размеченные понятия в тексте документа

Для каждой формулы, у которой есть связанные локальные переменные с привязанными понятиями, формируются набор связанных понятий и, как следствие, её семантическое наполнение. В Таблице 3 приведены примеры формул с распознанными понятиями по онтологии. Так, переменная, обозначающая понятие «Мера Лебега», была привязана к формуле, несмотря на то, что для этой переменной использовались различные символьные обозначения в разных документах.

В Таблице 4 приведены оценки точности, полноты и F-мера алгоритма связывания выделенных понятий с главной формулой. В эксперименте использовалась коллекция из 9 документов под авторством Р. Р. Кучарова, Ю. Х. Эшкабилова, Ю. Г. Никонорова, А. И. Парфенова, Н. М. Алиева, М. Э. Муминова, Л. Н. Ромакиной, Г. П. Арзикулова, В. Н. Берестовского, И. А. Зубаревой, И. С. Борисова,

С. Е. Хрущева, А. А. Боровкова, А. А. Могульского, Н. В. Васильевой, Н. В. Краснощёк⁵.

Таблица 3. Примеры формул с распознанными понятиями

Главная формула	Переменная	Понятие онтологии
$\int_{\Omega_j} \varphi_j(\xi) d\mu_j(\xi) = 0, \quad \int_{\Omega_j} \varphi_j^2(\xi) d\mu_j(\xi) = 1$	$\mu_j(\cdot)$	Мера Лебега
$t \mapsto \frac{n!}{\sqrt{\sum_{i=1}^n (x_i - x_0)^2}} \cdot \text{mes}(U(t)),$	$\text{mes}(U(t))$	Мера Лебега
$H = H_0 - (T_1 + T_2).$	H	Интегральный оператор
$H_2(\beta)\psi(y) = u(\beta)v(y)\psi(y) - \int_{\Omega_2} (\mu_0 + \mu\varphi_2(y)\varphi_2(t)) \psi(t) dt.$	$\{H_2(\beta)\}_{\beta \in \Omega_1}$	Семейство операторов
$P_H(x) = \sum_{k=0}^l \sum_{m=0}^{p_k-1} \sum_{s=0}^m \frac{f^{(m)}(z_k)}{(m-s)!(p_k-m-1)!} \frac{d^{m-s}}{dz^{m-s}} \left[\frac{(z-z_k)^{p_k}}{q(z)} \right] \Big _{z=z_k} \frac{q(x)}{(x-z_k)^{s+1}} \quad (7)$	P_H	Многочлен Эрмита

⁵ <http://mi.mathnet.ru/mt270>, <http://mi.mathnet.ru/mt271>, <http://mi.mathnet.ru/mt272>, <http://mi.mathnet.ru/mt273>, <http://mi.mathnet.ru/mt274>, <http://mi.mathnet.ru/mt275>, <http://mi.mathnet.ru/mt276>, <http://mi.mathnet.ru/mt277>, <http://mi.mathnet.ru/mt278>

Таблица 4. Оценки полноты и точности для алгоритма определения семантического наполнения формулы

Корректные распознавания (TP)	Разметка несуществующих понятий (FP)	Пропущенные понятия (FN)	Точность	Полнота	F-мера
190	16	18	0,92	0,91	0,92
120	8	56	0,94	0,68	0,79
136	48	36	0,74	0,79	0,76
192	18	48	0,91	0,80	0,85
495	120	50	0,80	0,91	0,85
340	12	45	0,97	0,88	0,92
441	32	10	0,93	0,98	0,95
400	64	9	0,86	0,98	0,92
450	54	24	0,89	0,95	0,92
Средние значения			0,89	0,88	0,88

ЗАКЛЮЧЕНИЕ

В статье представлен метод семантического аннотирования математических документов в формате PDF. Описан способ определения структуры документа с помощью разделения на блоки с текстом и главными формулами. Разработан метод связывания локальных переменных в тексте с главными формулами. С помощью метода разметки математических понятий в тексте формируется семантическое представление главных формул. Разработанный метод применяется для подготовки набора данных для семантического поисковика по формулам, который функционирует на цифровой платформе Lobachevskii-DML.

Будущие разработки связаны с расширением возможностей метода, в частности, для связывания многих переменных математической функции, а также разработки метода фильтрации понятий, обнаруженных в предложении, для увеличения точности метода аннотирования. Другим направлением для повышения универсальности метода можно считать планирование внедрения OCR-модуля

для распознавания текстов в отсканированных PDF, что особенно актуально для математических статей, изданных в доцифровую эпоху.

БЛАГОДАРНОСТИ

Исследование выполнено при поддержке Российского научного фонда, проект № 21-11-00105.

СПИСОК ЛИТЕРАТУРЫ

1. *Nevzorova O., Zhiltsov N., Zaikin D., Zhibrik O., Kirillovich A., Nevzorov V., Birialtsev E.* Bringing math to LOD: A semantic publishing platform prototype for scientific collections in mathematics // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2013. Vol. 8218 LNCS. No. 1. P. 379–394.
2. *Bertin M., Atanassova I.* Hybrid Approach for the Semantic Processing of Scientific Papers // *Semantic Publishing Challenge Track in 11th European Semantic Web Conference (ESWC 2014)*. 2014. P. 1–5.
3. *Ciancarini P., Di Iorio A., Nuzzolese A.G., Silvio P., Fabio V.* Semantic annotation of scholarly documents and citations // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2013. Vol. 8249 LNAI. P. 336–347.
4. *Ronzano F., Del Bosque G.C., Saggion H.* CEUR-WS proceedings: Towards the automatic generation of highly descriptive scholarly publishing linked datasets // *Communications in Computer and Information Science*. 2014. Vol. 475. P. 83–88.
5. *Ahmad R., Afzal M.T., Qadir M.A.* Information extraction from PDF sources based on rule-based system using integrated formats // *Communications in Computer and Information Science*. 2016. Vol. 641. P. 293–308.
6. *Greiner-Petter A., Youssef A., Ruas T., Miller, Bruce R., Schubotz M., Aizawa A., Gipp B.* Math-word embedding in math search and semantic extraction // *Scientometrics*. 2020. Vol. 125. No. 3. P. 3017–3046.
7. *Wolska M., Grigore M.* Symbol declarations in mathematical writing // *Proceedings of the 3rd Workshop on Digital Mathematics Libraries*. 2010. P. 119–127.
8. *Líška M., Sojka P., Ružička M., Mravec P.* Web interface and collection for

mathematical retrieval WebMlaS and MREC // DML 2011 – Towards a Digital Mathematics Library, Proceedings. 2011. P. 77–84.

9. Schubotz M., Greiner-Petter A., Scharpf P., Meuschke N., Cohl H.S., Gipp B. Improving the Representation and Conversion of Mathematical Formulae by Considering their Textual Context // Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. New York, NY, USA: ACM. 2018. P. 233–242.

10. Nevzorova O., Kirillovich A., Nevzorov V., Nikolaev K. The semantic context models of mathematical formulas in scientific papers // CEUR Workshop Proceedings. 2018. Vol. 2277. P. 33–40.

11. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A. OntoMath-PRO: Ontology of Mathematical Knowledge // Dokl. Math. 2022.
<https://doi.org/10.1134/S1064562422700016>

SEMANTIC ANNOTATION OF MATHEMATICAL FORMULAS IN PDF-DOCUMENTS

O. A. Nevzorova¹ [0000-0001-8116-9446], **K. S. Nikolaev**² [0000-0003-3204-238X]

¹⁻² *Kazan (Volga Region) Federal University, 35 Kremlyovskaya str., Kazan, 420008*

¹onevzoro@gmail.com, ²konnikolaeff@yandex.ru

Abstract

This article provides an overview of existing solutions for semantic analysis of mathematical documents, and also presents a method for automatic semantic analysis of documents in PDF format. This method searches for local variables in the text of the article, extracts their definitions and connects concepts with formulas. The advantage of the method over the existing ones is independence from the markup of the original PDF document, which expands the scope of the method. We provide estimates of recall, precision and F-measure for algorithms for finding variables and linking local variables with formulas. The resulting semantic markup of the document will be used to create a collection of documents suitable for the semantic formula search service, which is part of the set of services of the Lobachevskii-DML digital publishing system.

Keywords: *semantic analysis, PDF, document processing, scientific journals, Lobachevskii-DML.*

REFERENCES

1. *Nevzorova O., Zhiltsov N., Zaikin D., Zhibrik O., Kirillovich A., Nevzorov V., Birialtsev E.* Bringing math to LOD: A semantic publishing platform prototype for scientific collections in mathematics // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2013. Vol. 8218 LNCS. No. 1. P. 379–394.
2. *Bertin M., Atanassova I.* Hybrid Approach for the Semantic Processing of Scientific Papers // *Semantic Publishing Challenge Track in 11th European Semantic Web Conference (ESWC 2014)*. 2014. P. 1–5.
3. *Ciancarini P., Di Iorio A., Nuzzolese A.G., Silvio P., Fabio V.* Semantic annotation of scholarly documents and citations // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2013. Vol. 8249 LNAI. P. 336–347.
4. *Ronzano F., Del Bosque G.C., Saggion H.* CEUR-WS proceedings: Towards the automatic generation of highly descriptive scholarly publishing linked datasets // *Communications in Computer and Information Science*. 2014. Vol. 475. P. 83–88.
5. *Ahmad R., Afzal M.T., Qadir M.A.* Information extraction from PDF sources based on rule-based system using integrated formats // *Communications in Computer and Information Science*. 2016. Vol. 641. P. 293–308.
6. *Greiner-Petter A., Youssef A., Ruas T., Miller, Bruce R., Schubotz M., Aizawa A., Gipp B.* Math-word embedding in math search and semantic extraction // *Scientometrics*. 2020. Vol. 125. No. 3. P. 3017–3046.
7. *Wolska M., Grigore M.* Symbol declarations in mathematical writing // *Proceedings of the 3rd Workshop on Digital Mathematics Libraries*. 2010. P. 119–127.
8. *Líška M., Sojka P., Ružička M., Mravec P.* Web interface and collection for mathematical retrieval WebMlaS and MREC // *DML 2011 – Towards a Digital Mathematics Library, Proceedings*. 2011. P. 77–84.
9. *Schubotz M., Greiner-Petter A., Scharpf P., Meuschke N., Cohl H.S., Gipp B.*

Improving the Representation and Conversion of Mathematical Formulae by Considering their Textual Context // Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. New York, NY, USA: ACM. 2018. P. 233–242.

10. *Nezvorova O., Kirillovich A., Nezvorov V., Nikolaev K.* The semantic context models of mathematical formulas in scientific papers // CEUR Workshop Proceedings. 2018. Vol. 2277. P. 33–40.

11. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nezvorova O.A.* OntoMath-PRO: Ontology of Mathematical Knowledge // Dokl. Math. 2022.
<https://doi.org/10.1134/S1064562422700016>

СВЕДЕНИЯ ОБ АВТОРАХ



НЕВЗОРОВА Ольга Авенировна – доцент кафедры информационных систем Института вычислительной математики и информационных технологий Казанского федерального университета, к. т. н. Основные направления научных исследований: обработка естественного языка, искусственный интеллект.

Olga Avenirovna NEVZOROVA – Kazan Federal University, Institute of Computational Mathematics and Information Technologies, Associated Professor of the Department of Information System, PhD. Major fields of scientific research are Natural Language processing, artificial intelligence.

email: onevzoro@gmail.com

ORCID: 0000-0001-8116-9446



НИКОЛАЕВ Константин Сергеевич – ассистент кафедры системного анализа и информационных технологий Института Вычислительной математики и информационных технологий Казанского федерального университета. Основные направления научных исследований: обработка естественного языка, искусственный интеллект.

Konstantin Sergeevich NIKOLAEV – Assistant of the Department of System Analysis and Information Technologies of the Institute of Computational Mathematics and Information Technologies of Kazan Federal University. Major fields of scientific research are Natural Language processing, artificial intelligence.

email: konnikolaeff@yandex.ru

ORCID: 0000-0003-3204-238X

Материал поступил в редакцию 28 октября 2022 года

УДК 01; 002.53

РЕЙТИНГИ БИБЛИОГРАФИЧЕСКОЙ БАЗЫ И «БЕЛЫЕ СПИСКИ»

Т. А. Полилова^[0000-0003-4628-3205]

Институт прикладной математики им. М.В. Келдыша Российской академии наук, Миусская пл., 4, Москва, 125047

polilova@keldysh.ru

Аннотация

В настоящее время российские учреждения практически полностью отключены от западных информационных ресурсов и сервисов, связанных с изданием научных журналов. В таких условиях особую актуальность приобрела задача замещения ушедших сервисов, переориентация на отечественные научные журналы, российские библиотечные онлайн-ресурсы. В наиболее крупной библиографической базе eLibrary.ru, ориентированной на русскоязычные научные издания, собрана информация почти о 15 тысячах русскоязычных журналов. В eLibrary.ru действует аналитическая система «Российский индекс научного цитирования» (РИНЦ), обрабатывающая метаданные статей более чем 5 тысяч российских научных журналов. Насколько eLibrary.ru и РИНЦ готовы взять на себя функции национальной библиографической базы? По какой причине в российских организациях появляются «белые списки» журналов?

Основная проблема РИНЦ состоит в качестве построенных рейтингов научных журналов. Методики расчета рейтингов все прошедшие годы вызывали определенные критические замечания. В работе приводится пример построенного в РИНЦ рейтинга журналов из раздела «Математика». На первых позициях оказались журналы, малоизвестные в среде профессиональных математиков. Серьезные деформации в рейтингах РИНЦ подрывают доверие ученых к предлагаемым РИНЦ оценкам авторитетности российских журналов. Реакция некоторых вузов и научных организаций вполне ожидаемая: организации начинают вводить свои критерии успешности публикационной деятельности сотрудников, связанные с публикацией статей в журналах из так называемых «белых списков». Белый список журналов составляется, как правило, экспертными советами организации

прицельно по каждой дисциплине и научному направлению. При составлении белых списков наукометрические показатели могут учитываться, но они не являются преимущественным критерием для отбора журналов. Белые списки в настоящее время могут стать разумным дополнением рейтингов библиографических баз.

Ключевые слова: научная публикация, рейтинг журналов, тематическая классификация, импакт-фактор, мультидисциплинарность, библиографическая ссылка, белый список научных журналов

ВВЕДЕНИЕ

Недавно принятые радикальные санкции США и Евросоюза в отношении Российской Федерации в корне изменяют приоритеты публикационной стратегии российских ученых. Российские учреждения практически полностью отключены от западных информационных ресурсов, библиографических баз, многих сервисов, связанных с изданием научных журналов и статей. В таких условиях особую актуальность приобрела задача замещения ушедших сервисов, переориентация на отечественные научные журналы и российские библиотечные онлайн-ресурсы.

Наиболее крупной библиографической базой, ориентированной на русскоязычные научные издания, в настоящее время является eLibrary.ru [1], созданная в 2003 г. В этой библиографической базе собрана информация почти о 15 тысячах русскоязычных журналов, выходящих в настоящее время. На базе eLibrary.ru действует аналитическая система «Российский индекс научного цитирования» (РИНЦ) [2, 3], обрабатывающая метаданные статей более чем из 5 тысяч российских научных журналов. В eLibrary.ru зарегистрировано около 15 тысяч организаций, более 1 миллиона авторов статей. Насколько eLibrary.ru и РИНЦ готовы взять на себя функции национальной библиографической базы? По какой причине в российских организациях появляются «белые списки» журналов?

ФЕДЕРАЛЬНЫЙ СТАТУС

Любая крупная разработка в сфере IT требует устойчивых многолетних финансовых вложений. Трудно представить себе коммерческую компанию, готовую инвестировать в разработку электронных библиотечных и библиографических си-

стем. Программный продукт в этой сфере с трудом поддается коммерциализации, поскольку потребителем такого продукта в первую очередь являются бюджетные организации. Такого рода IT-разработки финансируются за счет государства. Электронная библиотека eLibrary на разных этапах получала бюджетную поддержку. Однако складывается впечатление, что многие методологические ошибки в eLibrary и РИНЦ проистекают от недостаточного финансирования. Небольшим коллективом исполнителей нелегко контролировать качество всех информационных ресурсов, накопленных к настоящему времени. Разработчикам приходится искать заинтересованных партнеров, пожелания и интересы которых, возможно, будут потом реализованы в конкретных сервисах, формулах расчета рейтингов, хитроумных библиометрических показателях.

Национальная библиографическая база (НББ) научных изданий должна стать независимым институтом сбора и обработки метаданных открытой научной продукции. В научных изданиях, представленных в НББ, публикуются в основном результаты научных исследований, выполняемых за счет средств федерального бюджета и средств региональных бюджетов. Текущая работа по контролю качества данных в НББ также должна финансироваться за счет бюджетных средств. Но не следует полностью исключать и создание в НББ платных сервисов с идеологией и функциональностью, открытыми и понятными научному сообществу.

Деятельность НББ должна быть поддержана на законодательном уровне. В РФ действует закон об обязательном экземпляре печатных документов и электронных копий печатных документов (оригинал-макетов). Как соотносятся база электронных копий печатных документов, формируемая законом об обязательном экземпляре, и будущая НББ? Понятно, что состав данных в этих двух базах не один и тот же. Поскольку закон об обязательном экземпляре должны выполнять и научные издания, все выходящие в печатном виде научные издания должны поставлять обязательные экземпляры выпусков в утвержденном порядке в ИТАР-ТАСС. Пока интернет-издания (электронные научные журналы) законом об обязательном экземпляре не охвачены. Но в НББ должны попадать метаданные выпусков и печатных, и онлайн-изданий. Для решения этой задачи потребуются разработка самостоятельных регламентов и процедур ввода метаданных.

Возможно, закон об обязательном экземпляре следует расширить, сделав обязательным представление в НББ метаданных статей научных изданий — и печатных, и онлайн-овых. В расширенном варианте этот закон будет способствовать снятию некоторых правовых коллизий в области авторского права.

Вопросы авторского права достаточно деликатные. В настоящее время можно встретить примеры судебных разбирательств, когда издатели и авторы статей требуют компенсаций за несанкционированное размещение в библиографических базах метаданных статей. Так, например, ООО «Медиамузыка» обратилось в арбитражный суд с иском к Российской академии наук и Центральному экономико-математическому институту РАН о нарушении исключительных прав на использование объектов интеллектуальной деятельности. Истец посчитал, что размещение в библиографической базе Соционет информации о нескольких произведениях, исключительные права на которые принадлежат истцу, незаконны. При этом имеется в виду размещение в Соционет аннотационных частей произведений, а также гиперссылок для доступа к полным текстам указанных произведений. Все эти гиперссылки начинаются с префиксной части адреса [https://socionet.ru/d/...](https://socionet.ru/d/)

В решении арбитражного суда по данному делу № А40-267527/21-12-1844 суд установил, что (цитируем): «На сайте Соционет нарушены права на доведение до всеобщего сведения названий и аннотационных частей семи произведений, а также на распространение семи спорных произведений, поскольку размещены гиперссылки [https://socionet.ru/d/...](https://socionet.ru/d/), создающие угрозу нарушения исключительного права» [4].

На наш взгляд, расширенный закон об обязательном экземпляре, предписывающий обязательную передачу в НББ метаданных статей, опубликованных в научных изданиях, создал бы правовую основу для функционирования НББ и возможного трансфера метаданных в другие научные базы.

В качестве еще одного примера разрешения спора на тему авторского права можно сослаться на статью А.В. Чернышова «Авторское право vs РФ», опубликованную на сайте электронного научного журнала «Медиамузыка» (эту статью можно найти, например, с помощью Яндекса). В статье сообщается, что Черны-

шов направлял в суд иск к Российской государственной библиотеке (РГБ) с требованием убрать из свободного доступа спорный сборник, в котором были обнаружены случаи плагиата. По мнению Чернышова, размещение сборника в РГБ нарушает права истца. Однако в этом иске Чернышову было отказано. Какова же причина такого решения? Суд отказал Чернышову в его иске, поскольку сборник попал в РГБ в соответствии с законом об обязательном экземпляре и по этой причине был открыт для читателей.

Означает ли это, что действия издателей, допустивших плагиат, не влекут юридической ответственности? Нет, не означает. Сам Чернышов сообщает, что он подал другой аналогичный иск к издателю-нарушителю авторских прав и потребовал изъять из продажи тираж. В этом случае суд принял положительное решение в пользу Чернышова.

СОТРУДНИЧЕСТВО МЕЖДУ РАН И НЭБ

В мае 2022 года РАН и eLibrary подписали новое соглашение о сотрудничестве. Оно стало продолжением и развитием предыдущего соглашения 2014 года, в рамках которого была создана коллекция научных изданий Russian Science Citation Index (RSCI) [5]. В RSCI вошли журналы, имеющие в РИНЦ высокие показатели цитируемости. В рамках нового соглашения стороны намерены продолжить и развивать сотрудничество по следующим направлениям:

— разработка инструментальных средств для агрегирования и учета научных изданий,

— развитие информационного массива журналов «Ядро Российского индекса научного цитирования (РИНЦ)» на основе журналов из RSCI, Web of Science и Scopus, а также совершенствование средств анализа, оценки и отбора в RSCI и «Ядро РИНЦ» изданий различного типа, в том числе сериальных изданий, сборников трудов научных мероприятий (семинаров, конференций, конгрессов). Предполагается, что в «Ядро РИНЦ» будут включены книги и другие типы научной литературы.

Хорошая новость состоит в том, что в рейтингах РИНЦ будут участвовать помимо журналов другие виды научных изданий, в частности труды конференций. В РИНЦ труды конференций являются пока второстепенными объектами, основное внимание уделяется периодическим изданиям — журналам. В то время как в

западных базах конференциям уделяется большое внимание. Так, например, в известной библиографической базе DBLP [6] число размещенных записей с метаданными трудов конференций (5898) превышает число записей с метаданными журналов (1805). Всего в этой базе индексируется 6 112 555 статей, зарегистрировано 2 999 744 автора. База DBLP специализируется в индексации изданий и статей по темам Computer Science. График пополнения этой базы, представленный на Рис. 1, показывает, что с каждым годом число индексируемых трудов конференций и семинаров (синий сектор) увеличивается быстрее, чем число индексируемых журналов (малиновый сектор).

Плохая новость состоит в том, что РАН по-прежнему ориентируется на созданную коллекцию RSCI, а также журналы, входящие в Web of Science и Scopus. Как мы помним, в коллекцию RSCI, включающую около 800 журналов, вошли научные журналы, имеющие, по мнению РАН, наилучшие показатели рейтинга РИНЦ. Однако к коллекции RSCI все прошедшие годы выдвигались претензии, касающиеся как методики определения «лучших» журналов, так и к составу коллекции RSCI [7]. Несмотря на декларации о широком демократическом обсуждении и экспертизе журналов, включаемых в RSCI, коллекция формировалась кулуарно, процедуры общественной экспертизы были непрозрачными. Не известны также параметры оценки журналов, ФИО и места работы экспертов, выставивших оценки журналам, полнота охвата экспертизой журналов РИНЦ. В RSCI, действительно, попали около 300 журналов с высокими показателями рейтингов РИНЦ. Эти известные и уважаемые в научном сообществе журналы издаются РАН, институтами РАН и ведущими вузами. Наряду с ними в RSCI попали журналы, показатели которых заметно уступали показателям журналов, не включенных в RSCI.

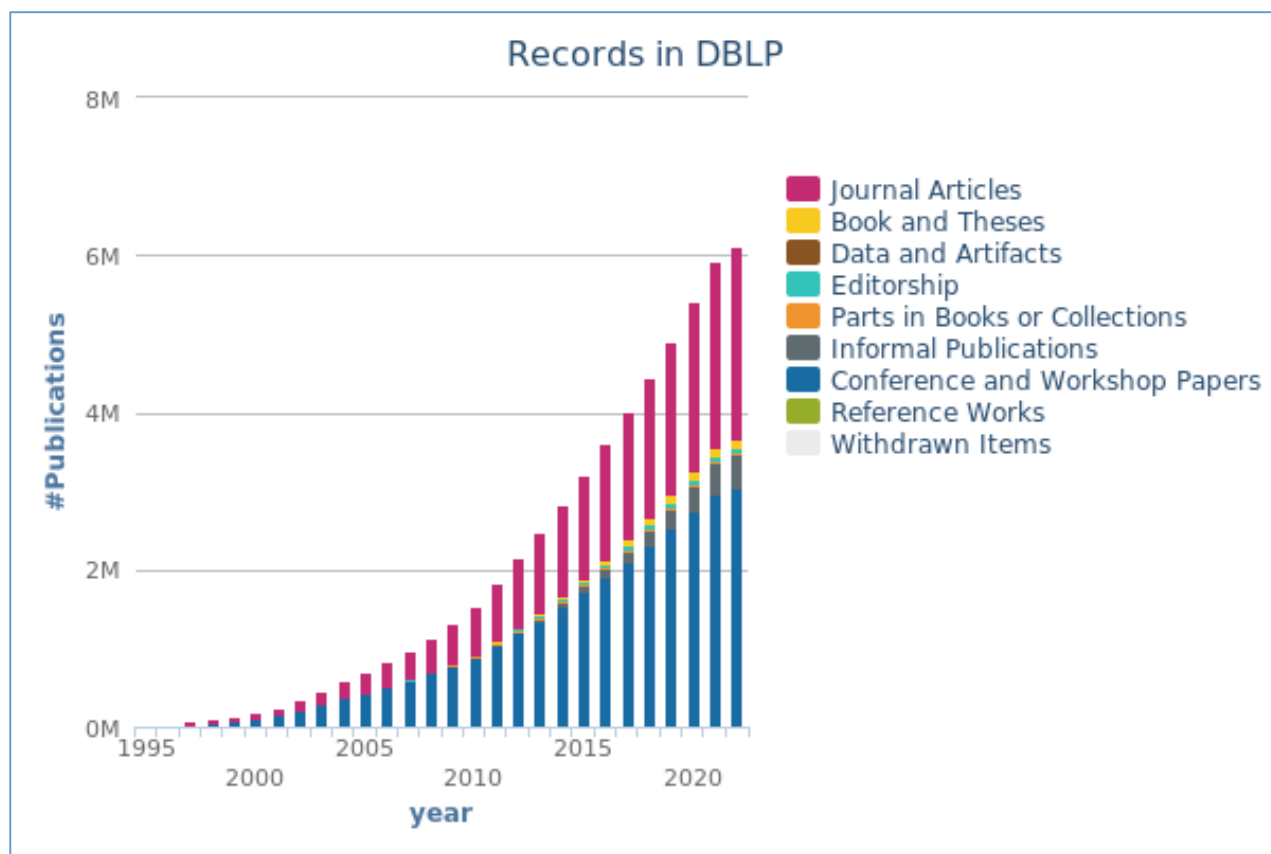


Рис. 1. Динамика наполняемости библиографической базы DBLP (<https://dblp.org/>).

Еще одна совместная с eLibrary неоднозначная инициатива РАН, демонстрирующая кулуарность принятия решений, — создание Академического рейтинга журналов RSCI [8]. Проект по созданию Академического рейтинга представил вице-президент РАН А.Р. Хохлов на заседании Президиума РАН 23 марта 2021 г. (Рис. 2) [9]. В отношении формулы Академического рейтинга были высказаны многочисленные критические замечания со стороны представителей редколлегий академических журналов, присутствующих на заседании [10]. Однако нет каких-либо сведений о том, как повлияли высказанные замечания на стратегию составления Академического рейтинга.

Следует отметить, что в формах РИНЦ для визуализации рейтингов журналов пользователями eLibrary.ru нет упоминаний об Академическом рейтинге. Вероятно, этот рейтинг был сконструирован исключительно для внутреннего употребления в РАН.

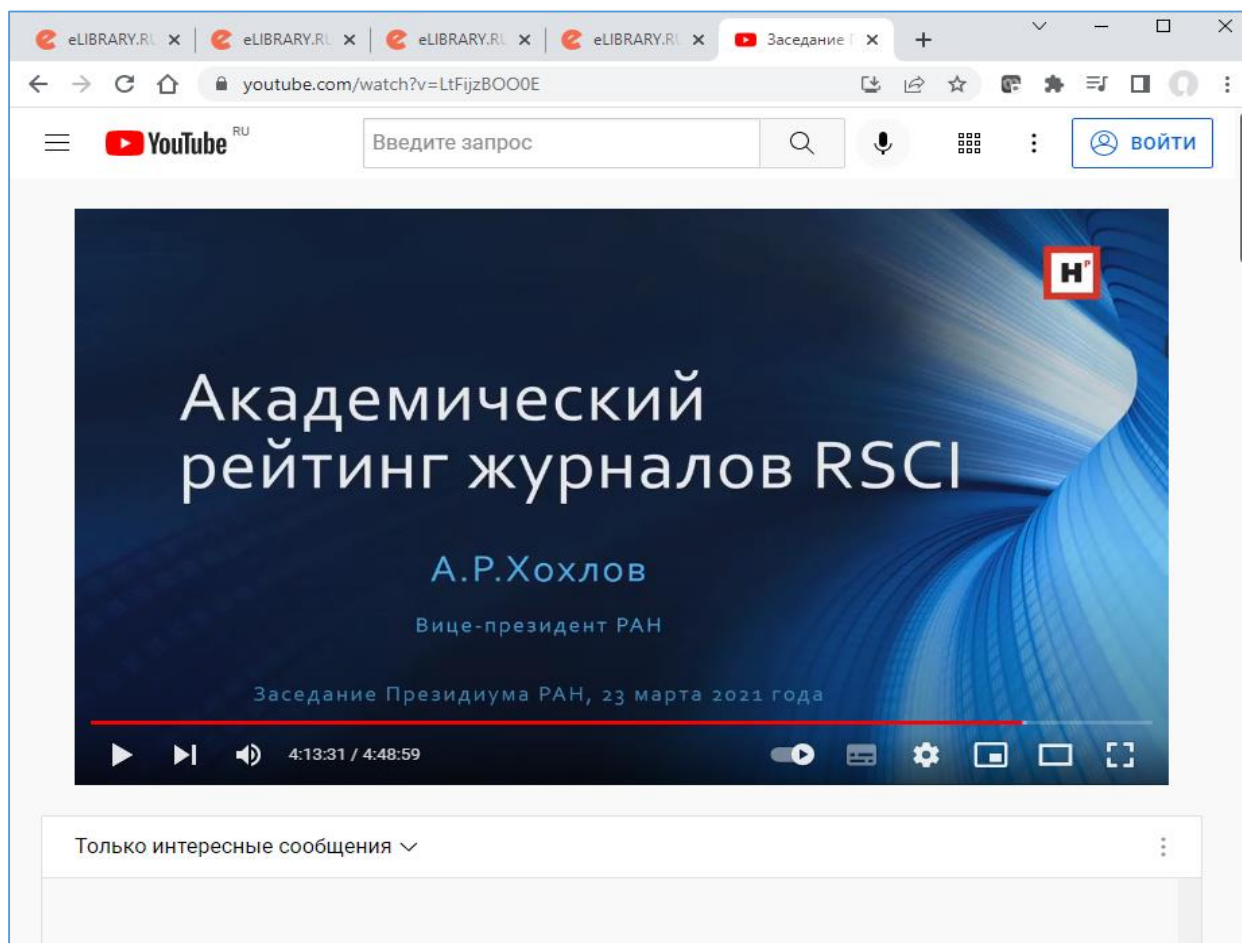


Рис. 2. Представление Академического рейтинга на заседании Президиума РАН 23 марта 2021 г.

По нашему мнению, желательно создать открытую площадку для обсуждения функциональности РИНЦ. Мнение широкой научной общественности, издательских структур академических институтов и вузов должно влиять на решения, принимаемые РИНЦ. Можно пойти дальше и создать постоянно действующую экспертно-аналитическую структуру, имеющую статус, независимый по отношению к РАН и РИНЦ, для контроля качества журналов и рейтингов РИНЦ.

Может ли РАН возглавлять такую экспертную структуру? Здесь можно усомниться в организационном потенциале РАН, породившей совместно с eLibrary.ru далеко не безупречную коллекцию RSCI и «Ядро РИНЦ». Сотрудники академических институтов вполне могли бы входить в такую структуру. Но наряду с ними экспертами должны стать представители вузов и других научных центров. Инфор-

мация о работе экспертно-аналитической структуры должна быть открытой и доступной пользователям eLibrary.ru и РИНЦ. Должны быть предусмотрены механизмы обратной связи, позволяющими научному сообществу влиять на принятие тех или иных важных решений в РИНЦ.

РЕЙТИНГИ РИНЦ

Основная проблема РИНЦ состоит в качестве построенных рейтингов научных журналов. Методики расчета рейтингов все прошедшие годы вызывали определенные критические замечания [10–14]. Много претензий высказано в адрес основного рейтинга РИНЦ — Science Index. Основу формулы рейтинга составляет пятилетний импакт-фактор, который традиционно используется и в других библиографических базах. Но наряду с этим показателем в формуле рейтинга Science Index участвует такой показатель, как индекс Херфиндаля [15], используемый экономистами при анализе процессов монополизации производственных процессов в обществе. У многих редакторов научных журналов возникают сомнения, насколько правомерно использовать в библиографических базах показатели, пришедшие из экономики.

В среде библиометристов сложилось мнение, что благодаря введению в формулу рейтинга показателя индекса Херфиндаля журналы, практикующие взаимное договорное цитирование, будут иметь слабые позиции. К сожалению, такие надежды разбиваются о подводные камни реальной жизни. Очевидно, что журналам, практикующим договорное цитирование, вполне по силам выстраивать особую редакционную политику, направленную на достижение подходящих значений индекса Херфиндаля.

Но и рейтинги РИНЦ по импакт-фактору, несмотря на простоту формулы вычисления показателя, также могут демонстрировать искажения, очевидные для специалистов в предметной области рейтинга.

Приведем следующий пример. На Рис. 3 представлены результаты рейтинга РИНЦ по пятилетнему импакт-фактору в разделе «Математика» за 2020 г.

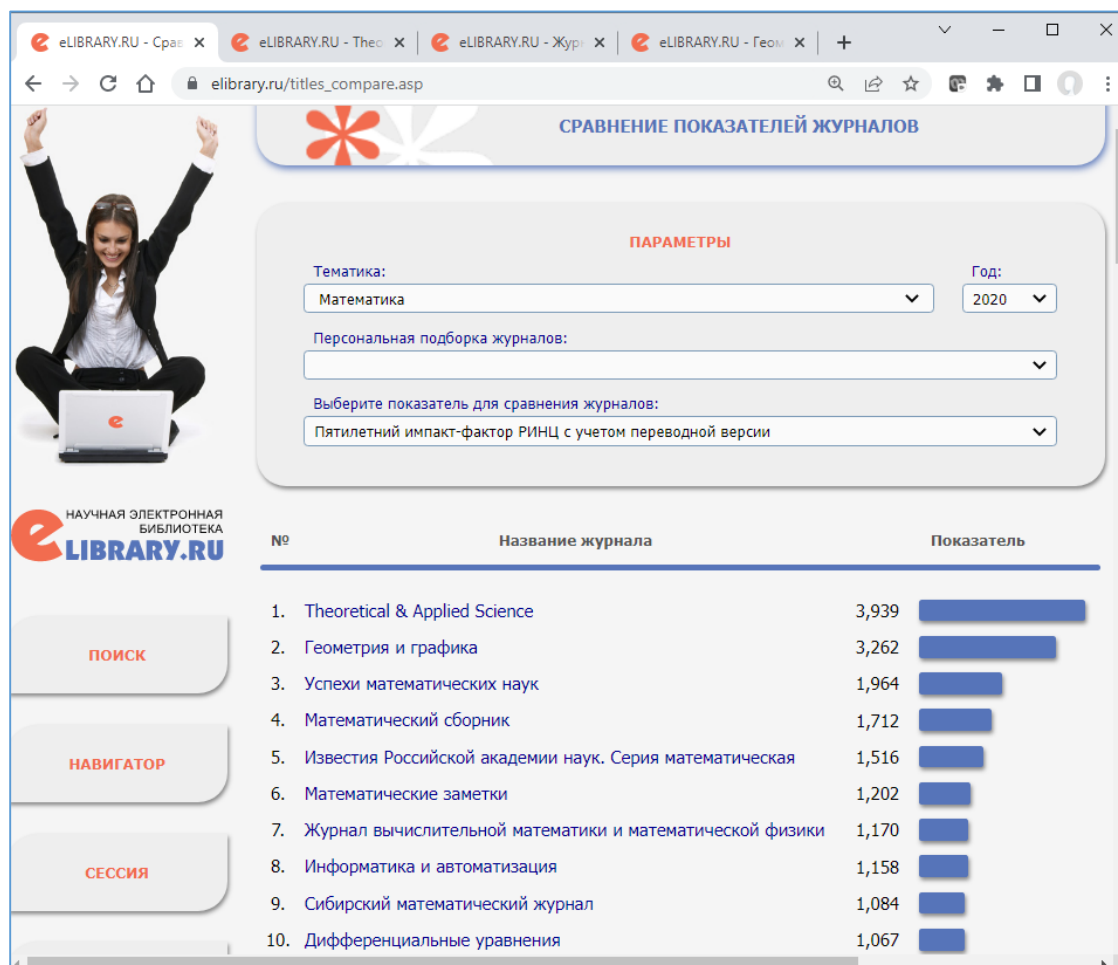


Рис. 3. Рейтинг РИНЦ 2020 года в разделе «Математика».

В Таблице 1 перечислены журналы, разместившиеся на первых 10 позициях рейтинга.

Таблица 1. Первые 10 позиций рейтинга РИНЦ по показателю «Пятилетний импакт-фактор».

№ в рейтинге	Название журнала	Показатель
1.	Theoretical & Applied Science	3,939
2.	Геометрия и графика	3,262

3.	Успехи математических наук	1,964
4.	Математический сборник	1,712
5.	Известия Российской академии наук. Серия математическая	1,516
6.	Математические заметки	1,202
7.	Журнал вычислительной математики и математической физики	1,170
8.	Информатика и автоматизация	1,158
9.	Сибирский математический журнал	1,084
10.	Дифференциальные уравнения	1,067

На первом месте рейтинга разместился журнал “Theoretical & Applied Science”, который по показателю импакт-фактора в несколько раз опередил классические математические журналы с непрекаемым авторитетом. Что представляет собой журнал “Theoretical & Applied Science”? На информационной странице журнала (https://www.elibrary.ru/title_about.asp?id=38848) читаем следующее:

“Theoretical & Applied Science” — Международный научный журнал, зарегистрированный во Франции и выходящий в формате Международных научно-практических конференций. Конференции проводятся ежемесячно – 30 числа в разных городах и странах. Все поданные авторами статьи в течение 1-го дня размещаются в интернете на сайте www.T-Science.org.

Столь оперативное размещение на сайте поступившей от авторов статьи (один день!) вызывает законный вопрос: является ли данный журнал рецензируемым? К сожалению, на информационной странице журнала никаких слов о рецензировании статей не приводится.

Журнал “Theoretical & Applied Science” является мультидисциплинарным, поскольку в нем заявлены 35 тематических рубрик верхнего уровня классификатора ГРНТИ. Почему этот журнал демонстрирует столь высокие показатели рейтинга? Если обратиться к странице со статистическими данными этого журнала https://elibrary.ru/title_profile.asp?id=38848, то можно увидеть, что количественные характеристики журнала весьма сомнительные.

В Таблице 2 представлены итоговые статистические показатели журнала, накопившиеся с года его основания (с 2013 г.).

Таблица 2. Итоговые статистические показатели журнала “Theoretical & Applied Science”.

Общее число выпусков журнала	109
Общее число статей из журнала	6075
Общее число статей с полными текстами	6074
Суммарное число цитирований журнала в РИНЦ	5381
Среднее число статей в выпуске	56
Число выпусков в год	12
Место в общем рейтинге SCIENCE INDEX за 2020 год	1911
Место в рейтинге SCIENCE INDEX за 2020 год в тематическом разделе «Математика»	65
Место в рейтинге по результатам общественной экспертизы	—
Средняя оценка по результатам общественной экспертизы	1,643
Число анкет с проставленной оценкой данному журналу	28(10,0%)

Обратим внимание, что общее число статей журнала в РИНЦ (строка 2) больше 6 тысяч. В Таблице 3 представлены более детальные сведения о журнале по годам (с 2011 по 2020), которые явно противоречат Таблице 1.

Таблица 3. Детальные сведения о журнале “Theoretical & Applied Science”

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Число статей в РИНЦ за год	0	0	1	1	0	0	0	0	131	1342
Общее число цитирований журнала в текущем году	0	1	0	25	52	74	91	103	151	736
Число выпусков журнала в РИНЦ за год	0	0	8	12	12	12	12	12	12	12
Число статей, опубликованных за предыдущие 5 лет	-	-	-	-	2	2	2	2	1	131
Число цитирований статей предыдущих 5 лет	0	0	0	17	34	65	69	78	117	516
Пятилетний импакт-фактор РИНЦ	-	-	-	-	-	-	-	-	-	3,939
Число полных текстов статей в elibrary.ru	-	-	121	214	173	50	225	416	-	-

Оказывается, РИНЦ рассчитал журналу пятилетний импакт-фактор, исходя из неполных данных (см. Таблицу 3, строку 1, выделенные цветом ячейки). Ошибка кроется в неправильном указании числа статей, опубликованных за период 2015–2019 гг., — суммарно 131. Именно этот показатель будет участвовать в формуле рейтинга в знаменателе дроби, при этом в числителе стоит число цитирований за указанный пятилетний период в 2020 г. — 516. На самом деле данные о числе опубликованных статей за этот период сильно занижены (за период 2015–2018 гг. число статей вообще не указано). Если исходить из данных Таблицы 2, то за 5 лет журнал должен был опубликовать примерно 672 статьи (в среднем журнал публикует по 56 статей за месяц). И если исходить из более правдоподобных данных, то импакт-фактор журнала был бы примерно 0,76.

Нетрудно заметить многочисленные нестыковки данных в приведенных Таблицах 2 и 3. Данным Таблицы 3 не соответствуют следующие позиции Таблицы 2:

- Общее число выпусков журнала;
- Общее число статей из журнала;
- Общее число статей с полными текстами;
- Суммарное число цитирований журнала в РИНЦ.

В Таблицах 2 и 3 приведены только числовые значения показателей. Но на странице статистических данных в eLibrary.ru общее число опубликованных статей и полученных цитирований являются гиперссылками. По ссылке «6075» можно перейти на список опубликованных статей, число которых должно быть 6075. Оказывается, что этот список насчитывает всего 3619 статей. РИНЦ дает сведения о тематической направленности опубликованных статей. Так, в разделе «Прикладная математика. Математическое моделирование. Теоретические исследования в математике» с учетом англоязычного варианта опубликовано всего 179 статей. Почему журнал участвует в рейтинге по направлению «Математика»?

Если перейти со страницы с показателями журнала по ссылке «Суммарное число цитирований журнала в РИНЦ» (строка 4 в Таблице 2), то вместо указанных 5381 цитирующих статей РИНЦ приводит список из 3323 статей, причем только 213 статей относятся к теме «Математика». Вопрос, заданный в предыдущем абзаце, можно повторить.

Приведенный пример показывает, что в РИНЦ не налажен содержательный контроль показателей журналов, участвующих в рейтингах. Понятно, что весьма трудоемко отслеживать качество представления информации в таблицах нескольких тысяч журналов. Но именно рейтинги, в первую очередь тематические, позволяют увидеть искажения.

Глядя на результаты рейтинга, любой специалист в области «Математика» тут же задаст вопрос: почему мультидисциплинарный журнал в форме заочных конференций попал на первую позицию в рейтинге по математике, оттеснив классические математические журналы? Если взглянуть на статистические данные журнала, то неизбежны следующие вопросы. Почему при вычислении показателей рейтингов нет проверки на полноту данных за требуемый период? Почему

данные в итоговой таблице (представлены в Таблице 2) на порядок отличаются от данных по годам (представлены в Таблице 3)?

Приведем еще один пример, связанный с проблемой построения адекватных рейтингов. На Рис. 4 представлены два графика: пятилетний импакт-фактор (голубой) и рейтинг Science Index (рыжий) в разделе «Математика» за 2020 г.

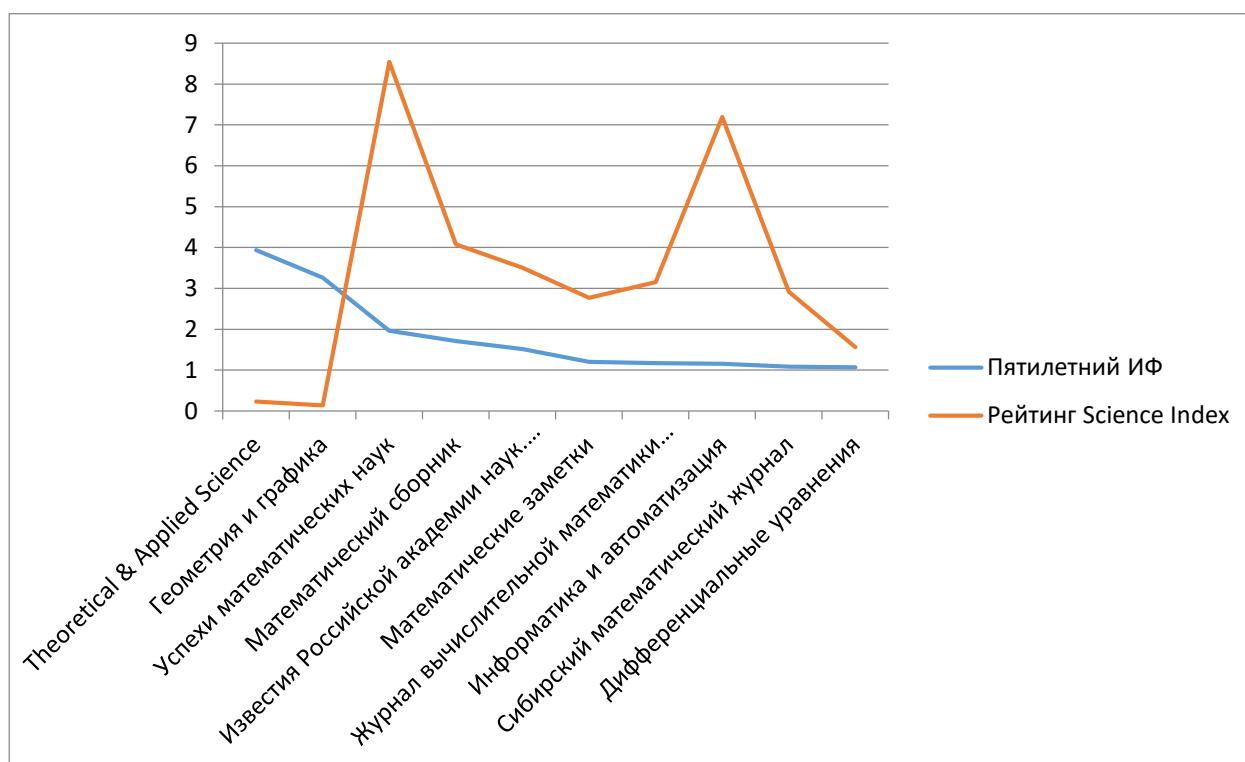


Рис. 4. Два графика РИНЦ в разделе «Математика» за 2020 г.

Как можно заметить, эти два графика мало коррелируются друг с другом, хотя оба этих рейтинга основаны на одном показателе — пятилетнем импакт-факторе. Какой из этих двух графиков больше отражает авторитет математических журналов? К сожалению, к обоим рейтингам есть претензии. В рейтинге по импакт-фактору (голубой график) непрофильные два журнала вышли на первые две позиции. Однако голубой график ожидаемо плавный. В рейтинге Science Index (рыжий график) один непрофильный журнал «Информатика и автоматизация» вышел на второе место. Но пики на графике настораживают. Скорее всего, необычные выбросы провоцирует индекс Херфиндаля [15] в формуле рейтинга Science Index. Хотя неочевидно, что журналы, дающие выбросы на графике рей-

тинга Science Index, демонстрируют некие уникальные качественные характеристики, выводящие эти журналы на более высокие позиции рейтинга Science Index. Скорее всего, редакции таких журналов предпринимают особые усилия для получения нужных значений индекса Херфиндаля.

Вернемся к рассмотрению Таблицы 1, демонстрирующей первые 10 позиций рейтинга РИНЦ по показателю «Пятилетний импакт-фактор». Помимо лидирующего положения в рейтинге журнала “Theoretical & Applied Science”, вызывает удивление попадание на вторую позицию рейтинга по математике журнала «Геометрия и графика». В работах [11, 12] уже высказывались замечания по методикам определения референтных групп журналов, участвующих в тематическом рейтинге по математике. На наш взгляд, журнал «Геометрия и графика», занявший вторую позицию в тематическом рейтинге в разделе «Математика», не должен участвовать в рейтинге в этом тематическом направлении. Этот журнал заявлен в Перечне ВАК по следующим специальностям:

- 05.01.01 Инженерная геометрия и компьютерная графика (технические науки) (с 28.12.2018);
- 5.8.2. Теория и методика обучения и воспитания (по областям и уровням образования) (педагогические науки);
- 5.8.7. Методология и технология профессионального образования (педагогические науки) (с 01.02.2022).

В анкете журнала «Геометрия и графика» указаны следующие тематические рубрики.

РУБРИКИ ГРНТИ:

143500. Высшее профессиональное образование. Педагогика высшей профессиональной школы;

272100. Геометрия;

281700. Теория моделирования.

РУБРИКИ OECD:

101. Mathematics;

503. Educational sciences.

По-видимому, журнал отнесен к направлению «Математика» на основании заявленных журналом рубрик ГРНТИ и OECD. Хотя такое решение явно противоречит специальностям номенклатуры ВАК, указанным журналом: все три специальности не относятся к классической математике, при этом две относятся к направлению «Педагогические науки».

Гармонизировать статический классификатор ГРНТИ, номенклатуру специальностей ВАК, рубрикатор OECD для решения задач библиографической базы представляется делом сложным и неблагодарным. Если бы РИНЦ использовал современный, динамически формирующийся тематический классификатор, основанный на библиографических цитированиях [16, 17], то проблема отбора референтных групп журналов в тематических рейтингах решалась бы аккуратнее – журнал «Геометрия и графика» вряд ли попал бы в кластер классических математических журналов.

Отметим, что РИНЦ пока слабо использует тематики номенклатуры ВАК в процедурах построения рейтингов. Хотя известно, что для многих ученых основным мотивом написания статей является подготовка диссертаций по определенной специальности. В соответствии с требованиями ВАК результаты диссертационных исследований должны публиковаться в журналах из Перечня ВАК по соответствующей специальности из номенклатуры ВАК. Тем самым статьи на содержательном уровне естественным образом привязываются авторами к одной или нескольким специальностям ВАК. При этом качество такой привязки проверяется многочисленными экспертами. Сначала – членами диссертационных советов, где проводится защита диссертации. Члены диссертационного совета должны проверить содержание и полноту опубликования результатов диссертации в печати. Далее в процессе оценки диссертации в ВАК полнота опубликования результатов в профильных журналах, привязанных к номенклатуре ВАК, оценивается членами экспертных советов ВАК.

Как показывает недавно опубликованный Перечень рецензируемых изданий, ВАК перешла к более строгому учету соответствия тематики журналов специальностям номенклатуры ВАК. При регистрации журнала для включения в Перечень ВАК редакция журнала не голословно указывает специальность ВАК – редакция должна привести несколько статей по указанной специальности из выпусков

издания за прошедший полугодовой период. Кроме того, редакция указывает, кто именно из членов редколлегии с учеными степенями имеет соответствующие компетенции по указанной специальности номенклатуры ВАК.

Можно предположить, что при защите диссертации по определенной специальности соискатель должен будет иметь публикации в журналах из Перечня именно по выбранной им специальности из номенклатуры ВАК. Вообще говоря, это новый поворот требований к соискателю: в действующем Положении о защите диссертаций явно не указана такая привязка, но она вполне ожидаема.

Почему бы в eLibrary не привязывать журналы к специальностям ВАК, отталкиваясь от текущего состояния Перечня? Опора на Перечень ВАК позволяет более аккуратно решить проблему выбора референтных групп журналов для тематических рейтингов. Разумеется, для этого необходимо сотрудничество между eLibrary и ВАК. Желательно оперативно передавать в eLibrary метаданные журналов из Перечня ВАК в автоматизированном режиме. В этом случае РИНЦ мог бы строить рейтинги журналов из актуального Перечня для каждой специальности из номенклатуры ВАК. Такие рейтинги могли бы стать хорошим ориентиром для соискателей ученых степеней, выбирающих журнал для опубликования своих статей. Сейчас, к сожалению, удобных средств для выбора журнала по конкретной специальности номенклатуры ВАК соискателю не предоставляется ни на сайте Минобрнауки, ни на сайте ВАК, ни в eLibrary.

О БЕЛЫХ СПИСКАХ НАУЧНЫХ ЖУРНАЛОВ

Серьезные деформации в рейтингах РИНЦ подрывают доверие ученых к предлагаемым РИНЦ оценкам авторитетности российских журналов. Реакция некоторых вузов и научных организаций вполне ожидаемая: организации начинают вводить свои критерии успешности публикационной деятельности сотрудников, связанные с публикацией статей в журналах из так называемых «белых списков». Белый список журналов составляется, как правило, экспертными советами организации прицельно по каждой дисциплине и научному направлению. При составлении белых списков наукометрические показатели могут учитываться, но они не являются преимущественным критерием для отбора журналов. Эксперты хорошо понимают ценность и авторитет того или иного научного журнала в своей области,

профессионализм рецензентов, качество публикуемых материалов и аудиторию читателей.

Высшая школа экономики (НИУ ВШЭ) на своем сайте ввела раздел «Новая система начисления академических надбавок» [18] (Рис. 5).

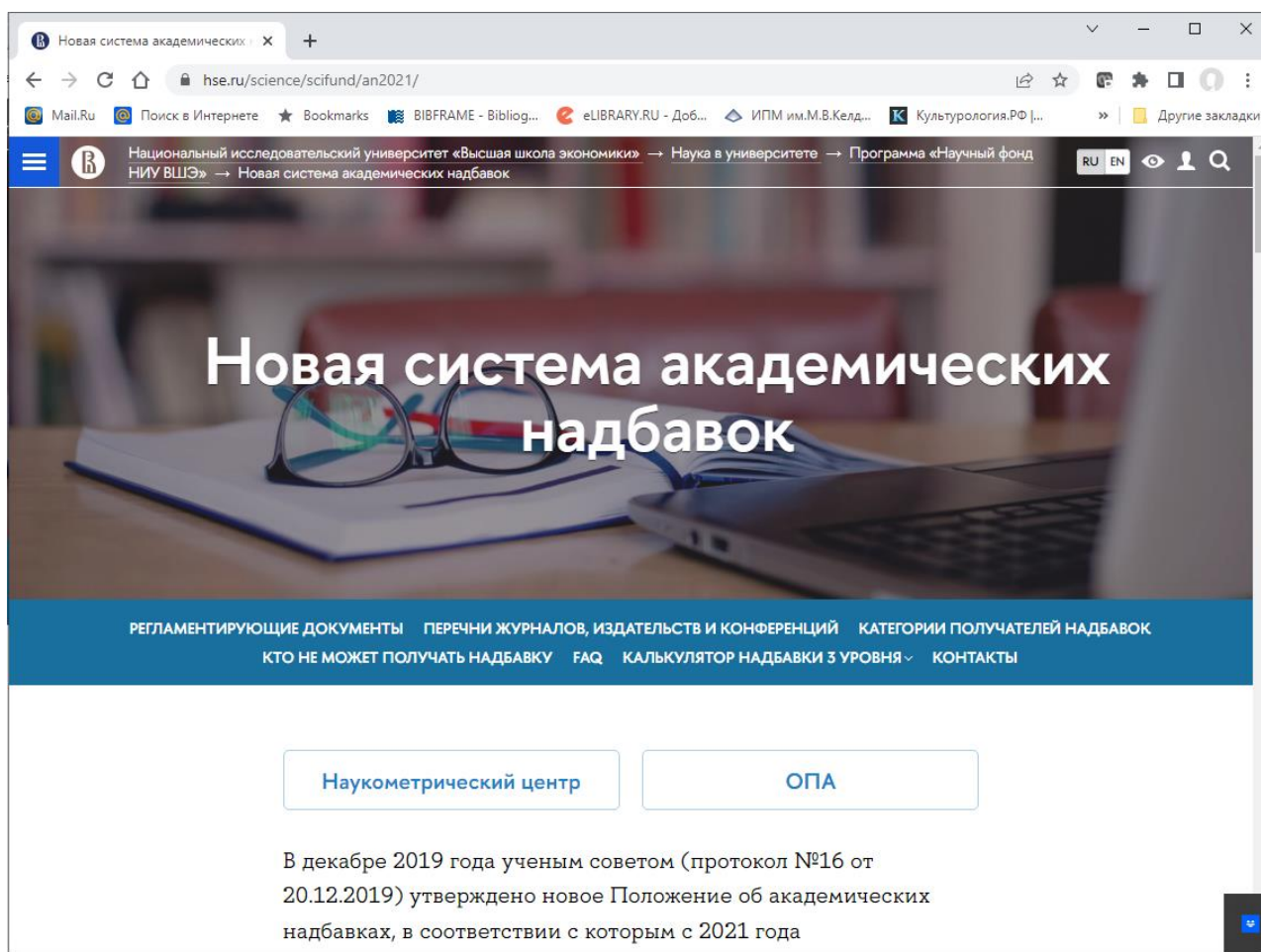


Рис. 5. Новая система академических надбавок НИУ ВШЭ.

В системе надбавок НИУ ВШЭ введено несколько белых списков для разных видов академических поощрений. Используются списки с буквенными обозначениями А+, А, В, С, D и Z. При этом список А+ включает наиболее престижные мировые издания. Журналы из этого списка учитываются при назначении «мега-надбавок». Доступ к спискам А и В закрыт для пользователей интернета (вход только по контролируемым реквизитам). В список С включаются журналы, индексируемые в Web of Science и Scopus. В списке D содержится весьма представительный список иностранных и русскоязычных журналов. Список Z по существу

является «черным списком», публикации в журналах из этого списка не учитываются в процедурах надбавок НИУ ВШЭ.

Действующие белые списки формируются с участием отраслевых рабочих групп НИУ ВШЭ и обновляются один раз в три года. Списки публикуются на сайте Научометрического центра НИУ ВШЭ [19]. Центр также может по своему усмотрению проводить экспертизу журналов. Если по результатам внеочередной экспертизы Центра принято решение об исключении журнала из белых списков (без включения в список Z), решение вступает в силу при ближайшем обновлении списков. В случае внесения журнала в список Z, он исключается из белых списков с момента принятия такого решения.

Свой белый список подготовила и Российская академия наук [20] (Рис. 6).

«Белый список» РАН – это перечень научных журналов, который предполагается использовать для оценки результативности научных организаций и расчета комплексного бала публикационной результативности (КБПР) [21]. «Белый список» РАН подготовлен межведомственной рабочей группой при активном участии академика А.Р. Хохлова. Сообщается, что в состав рабочей группы вошли, помимо представителей РАН, сотрудники Российского центра научной информации (РЦНИ), крупнейших отечественных вузов и научных организаций. «Белый список» РАН включает западные научные журналы, индексируемые в Web of Science Core Collection и Scopus, и российские журналы (в объеме коллекции RSCI).

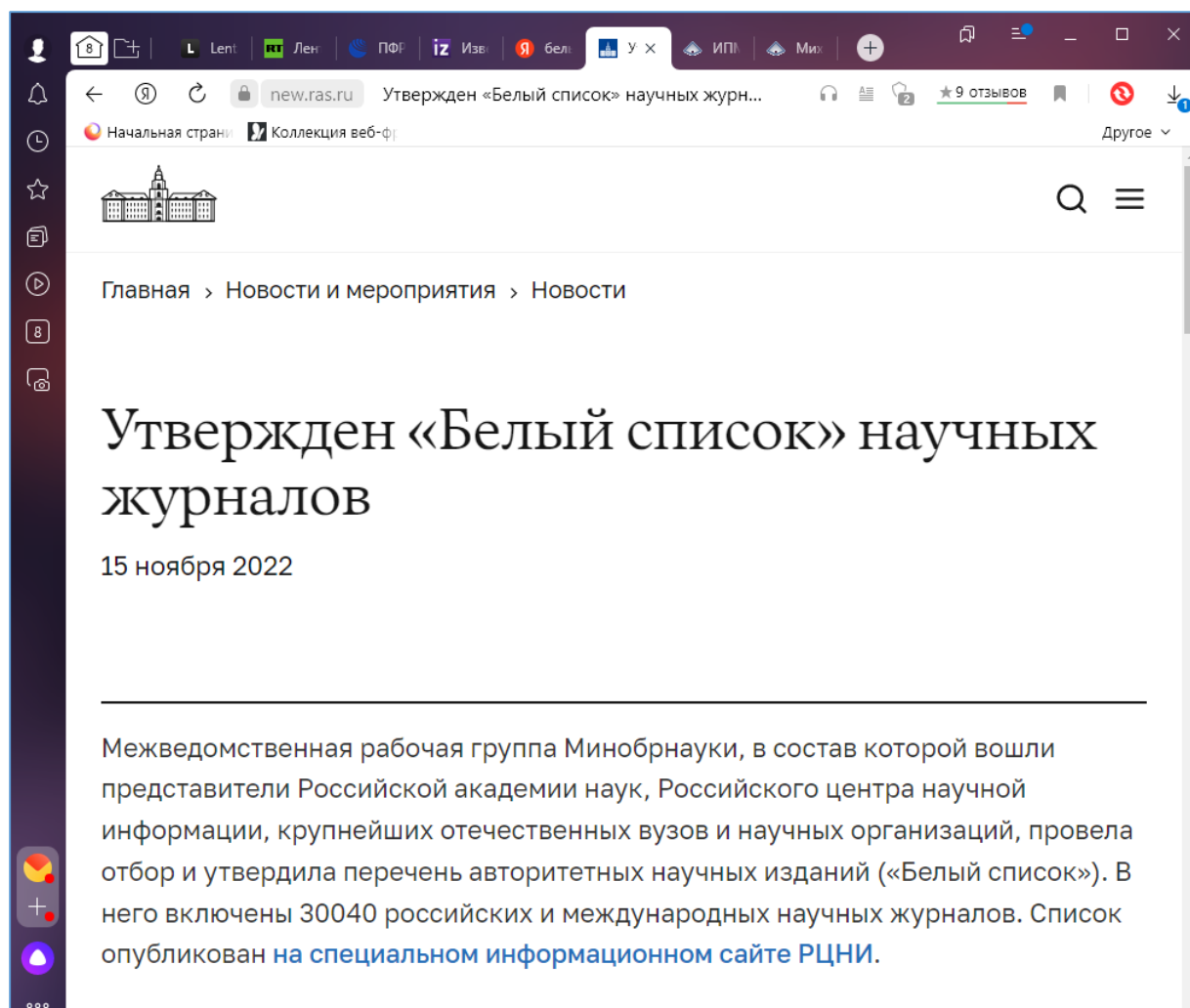


Рис. 6. Сообщение об утверждении «Белого списка» на сайте РАН.

«Белый список» РАН опубликован на специальном информационном сайте РЦНИ [22] (Рис. 7). Разработчики сайта обещают ежемесячно обновлять список и дополнять новыми полями карточки журналов. По словам А.Р. Хохлова, причиной изменения «Белого списка» РАН могут стать как включение лучших российских журналов, которые пока не входят в RSCI, так и исключение «хищных» журналов, которые все еще индексируются в зарубежных базах данных.

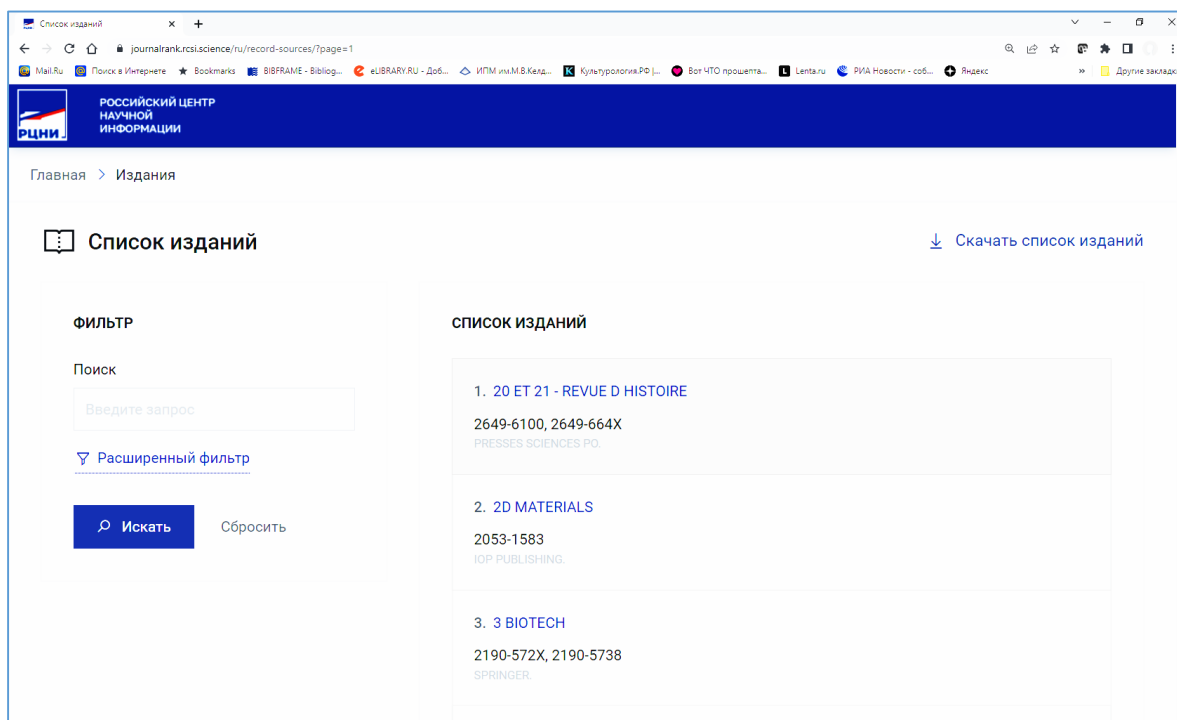


Рис. 7. «Белый список» РАН на сайте Российского центра научной информации.

Обратим внимание на пропорцию российских и западных журналов в «Белом списке» РАН. Всего в него входят около 30 тысяч наименований журналов. Напомним, что коллекция RSCI содержит около 800 российских журналов. Таким образом, разработчики этого «Белого списка» с очевидностью стимулируют российских авторов публиковаться в западных журналах. Однако остаются без ответа следующие вопросы. Насколько открыты западные журналы для российских авторов? Все ли журналы, включенные в столь амбициозный список, отвечают высоким требованиям качества?

Приходится констатировать, что «Белый список» РАН — это попытка обойти Постановление правительства РФ, принятое в марте 2022 г., о приостановке учета с повышающими коэффициентами публикаций в журналах, индексируемых в библиографических базах Web of Science и Scopus. Как отмечает министр высшего образования В.Н. Фальков, целью Постановления является необходимость «пересмыслить работу на данном этапе и обеспечить поддержку российских научных изданий, а также снизить удельный вес библиометрических и наукометрических показателей в оценке научных коллективов» [23]. РАН через межведомственную

рабочую группу фактически пытается сохранить ориентацию на западные журналы и библиометрические инструменты, несмотря на изменившиеся социальные реалии.

Заметим, что здесь речь не идет о борьбе с публикациями российских ученых на английском языке. Как отмечается в статье ректора МГИМО А.В. Торкунова [24] на сайте Российского совета по международным делам, английский язык давно перестал быть исключительно языком «англосаксов». Судя по количеству публикаций в международных библиографических базах, на английском языке издается значительная часть научных трудов в Китае и Индии, в странах Африки и Латинской Америки. Англоязычные публикации позволяют национальным научным школам повысить престиж и обеспечить продвижения своих результатов на глобальном уровне.

В этой же статье [24] отмечается, что предназначение и ценность для России библиографических баз Web of Science и Scopus осознаются и трактуются неверно. По существу, это всего лишь хорошо структурированный архив научной периодики, поддерживающий единые правила отбора и подготовки научных работ. Хотя не исключено, что есть и другие, не менее достойные архивы научных публикаций, в частности, библиографические базы журналов открытого доступа, национальные индексы стран Азии, Африки и Латинской Америки.

Угрозы ограничения доступа российских работ в западные базы дают дополнительный повод внедрить собственные меры поддержки институциональных основ научной работы в России. Автор статьи [24] предлагает в частности рекомендовать научным и образовательным организациям сформировать в целях внутренней оценки результативности научных исследований собственный белый список журналов из лучших (не обязательно лидеров по наукометрическим параметрам) международных журналов, не замеченных в дискриминации российских авторов.

В настоящее время обновляется еще один белый список — Перечень ВАК. Готовится соответствующая реформа требований к научным публикациям соискателей ученых степеней. Сообщается [25], что одной из причин реформы стало ограничение в международных базах Web of Science и Scopus индексирования

российских журналов. В марте 2022 г. Постановлением правительства РФ требование о публикации в журналах из этих библиографических баз было отменено. ВАК намерена составить новый список библиографических западных баз, в которых индексируются российские журналы. Попадание журнала в такие базы будет означать удовлетворение требований ВАК в отношении возможности публиковать результаты диссертационных исследований.

ВАК также вводит категорирование российских журналов в перечне ВАК, который содержит примерно 2700 журналов [26].

Предлагается ранжировать Перечень по группам по аналогии с квартилями Web of Science и Scopus Q1 Q2 Q3 Q4. Более конкретно — предлагается ввести группы К1 (квартиль Q1), К2 (объединяет квартили Q2 и Q3) и К3 (в которую попадают журналы из квартиля Q4). Участие в ранжировании Перечня примут экспертные советы ВАК. Ранжирование будет проводиться по нескольким показателям с учетом показателей цитируемости.

По-видимому, мы наблюдаем определенное соперничество между двумя структурами — РАН и ВАК. Тем не менее, хотелось бы, чтобы сильные академические журналы в результате не оказались за бортом Перечня ВАК. Как пояснил председатель ВАК В.М. Филиппов [26], около 90% журналов из RSCI уже входят в Перечень ВАК.

ЗАКЛЮЧЕНИЕ

Объективно электронная библиотека eLibrary.ru имеет серьезные шансы стать Национальной библиографической базой. Однако eLibrary.ru и РИНЦ должны наладить регулярный контроль полноты и достоверности статистических и библиометрических показателей журналов.

Можно рекомендовать РИНЦ существенно перестроить методики построения рейтингов. В основе тематической классификации журналов сейчас лежит весьма устаревший статический классификатор ГРНТИ. При выделении референтных групп журналов РИНЦ мог бы разработать и использовать современные алгоритмические классификаторы, основанные на кластерном анализе публикаций,

связанных библиографическим цитированием. Алгоритмы анализа близости журналов по библиографическим ссылкам помогли бы РИНЦ строить более адекватные тематические рейтинги в референтных группах журналов.

Появление белых списков свидетельствует о попытке переосмыслить значение библиометрических и наукометрических показателей при оценке научных журналов и снизить удельный вес этих показателей в оценке эффективности научной деятельности. Белые списки журналов составляются экспертными сообществами для определения авторитетных журналов в конкретных научных областях. Эти списки далеко не всегда совпадают со списком высокорейтинговых журналов, имеющих высокие библиометрические показатели, но могут стать разумным дополнением рейтингов библиографических баз.

СПИСОК ЛИТЕРАТУРЫ

1. Научная электронная библиотека eLibrary.ru. <https://www.elibrary.ru/>
 2. Российский индекс научного цитирования.
https://www.elibrary.ru/project_risc.asp
 3. *Анна Уфимцева*. Национальная система оценки науки может быть построена на данных РИНЦ. Глава Elibrary.ru рассказал, сможет ли Россия отказаться от международных баз научных журналов // Газета.ru. 14.03.2022. <https://www.gazeta.ru/science/2022/03/14/14624287.shtml>
 4. Решение арбитражного суда РФ от 22.02.2022 г. по делу № А40-267527/21-12-1844. https://kad.arbitr.ru/Document/Pdf/9d4cf831-4442-4cb5-9519-b4686ab972e8/4a5e8937-4412-4657-a087-5a901de76d35/A40-267527-2021_20220222_Reshenija_i_postanovlenija.pdf
 5. О проекте Russian Science Citation Index.
https://www.elibrary.ru/rsci_about.asp
 6. DBLP computer science bibliography. <https://dblp.org/>
 7. *Мазов Н.А., Гуреев В.Н., Каленов Н.Е.* Некоторые оценки списка журналов Russian Science Citation Index // Вестник РАН. 2018. Т. 88, № 4. С. 322–332. <https://doi.org/10.7868/S0869587318040047>
 8. Академический рейтинг журналов RSCI.
<http://www.ras.ru/news/shownews.aspx?id=6ddc26ee-044b-4cc9-99a0-96670f39ac13>
-

9. Представление Академического рейтинга на заседании Президиума РАН 23 марта 2021 г. Выступление А.Р. Хохлова.

<https://www.youtube.com/watch?v=LtFijzBOO0E> (4:12:45)

10. *Полилова Т.А.* РАН создает академический рейтинг научных журналов. // Научный сервис в сети Интернет: труды XXII Всероссийской научной конференции (20–23 сентября 2021 г., онлайн). М.: ИПМ им. М.В.Келдыша, 2021. С. 279–296.

<https://doi.org/10.20948/abrau-2021-32>,

<https://keldysh.ru/abrau/2021/theses/32.pdf>

11. *Полилова Т.А.* Рейтинги журналов в РИНЦ как инструменты анализа и влияния // Препринты ИПМ им. М.В. Келдыша. 2021. № 40. 35 с.

<https://library.keldysh.ru/preprint.asp?id=2021-40>;

<https://doi.org/10.20948/prepr-2021-40>

12. *Полилова Т.А., Горбунов-Посадов М.М.* Рейтинг журнала в библиографической базе // Электронные библиотеки. 2021. Т. 24, № 6. С. 1060–1089.

<https://doi.org/10.26907/1562-5419-2021-24-6-1060-1089>

13. *Gorbunov-Posadov M.V., Polilova T.A.* The Problem of Selecting a Reference Group for the Thematic Rating of Scientific Journals // CEUR Workshop Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). 2022. Vol. 3066. P. 90–100. <http://ceur-ws.org/Vol-3066/paper9.pdf>

14. *Полилова Т.А.* Препринты ИПМ им. М.В. Келдыша в диаграммах Science Space // Препринты ИПМ им. М.В. Келдыша. 2022. № 27. 38 с.

<https://doi.org/10.20948/prepr-2022-27>;

<https://library.keldysh.ru/preprint.asp?id=2022-27>

15. Herfindahl–Hirschman Index. From Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Herfindahl%E2%80%93Hirschman_Index

16. *Potter I.* Introducing Citation Topics in InCites – Clarivate. Dec 3, 2020. URL: <https://clarivate.com/blog/introducing-citation-topics/>

17. *Касьянов П.* Классификатор предметных областей Citation Topics в системе InCites. <https://youtu.be/58g6dQyJCW8>

18. Раздел «Новая система академических надбавок» на сайте НИУ ВШЭ.

<https://www.hse.ru/science/scifund/an2021/>

19. Научомеритрический центр НИУ ВШЭ. <https://scientometrics.hse.ru/> .
20. Утвержден «Белый список» научных журналов. 2022.
<https://new.ras.ru/activities/news/utverzhden-belyy-spisok-nauchnykh-zhurnalov/>
21. *Беляева С.* Того и ждали: «Белый список» научных журналов никого не удивил // Поиск. №47 (1745) от 18 ноября 2022. С. 4.
https://poisknews.ru/wp-content/uploads/2022/11/poisk_47_20221118.pdf
22. «Белый список» РАН на сайте Российского центра научной информации.
<https://journalrank.rcsi.science/ru/record-sources/>
23. Власти ввели мораторий на учет научных статей в международных изданиях // РБК. 21 марта 2022 г.
<https://www.rbc.ru/society/21/03/2022/6238915c9a794774a68b0245>
24. *Торкунов А.* Общественные науки в России: журналы и индексы. 2022.
<https://russiancouncil.ru/analytics-and-comments/comments/obshchestvennye-nauki-v-rossii-zhurnaly-i-indeksy/>
25. *Курилова А., Мишина В.* ВАК готовит реформу требований к научным публикациям соискателей степеней // Ведомости. 29 ноября 2022.
<https://www.vedomosti.ru/society/articles/2022/07/25/932990-vak-reformu-trebovanii-publikatsiyam>
26. ВАК поделил научные журналы на три категории. 16.12.22.
<https://poisknews.ru/science-politic/vak-podelil-nauchnye-zhurnaly-na-tri-kategorii/>
-

BIBLIOGRAPHIC DATABASE RATINGS AND WHITE LISTS

T. A. Polilova^[0000-0003-4628-3205]

Keldysh Institute of Applied Mathematics, Miusskaya sq., 4, Moscow, 125047, Russia

polilova@keldysh.ru

Abstract

Currently, Russian institutions are almost completely disconnected from Western information resources and services related to the publication of scientific journals.

In such conditions, the task of replacing the departed services, reorientation to domestic scientific journals, Russian online library resources has become particularly actual. In the largest bibliographic database the eLibrary.ru, focused on Russian-language scientific publications, collected information about almost 15 thousand Russian-language journals. In the eLibrary.ru there is an analytical system "Russian Science Citation Index" that processes metadata of articles from more than 5 thousand Russian scientific journals. Is the eLibrary.ru ready to serve as a national bibliographic database? For what reason "white lists" of journals appear in Russian organizations?

The main problem of the RSCI is the quality of the constructed ratings of scientific journals. The methods of calculating ratings over the past years have caused certain criticisms. The paper provides an example of a rating of journals from the section "Mathematics" built in the RSCI. Journals that are little known among professional mathematicians were in the first positions. Serious deformations in the ratings of the eLibrary.ru undermine the confidence of scientists in the assessments of the credibility of Russian journals proposed by the eLibrary.ru. The reaction of some universities and scientific organizations is quite expected: organizations are beginning to introduce their own criteria for the success of the publication activities of employees associated with the publication of articles in journals from the so-called "white lists". The white list of journals is compiled, as a rule, by the expert councils of the organization specifically for each discipline and scientific direction. Scientometric indicators may be taken into account when compiling white lists, but they are not the primary criterion for the selection of journals. White lists can now become a reasonable addition to the ratings of bibliographic databases.

Keywords: *scientific publication, rating of journals, thematic classification, impact factor, multidisciplinary, bibliographic reference, white list of scientific journals*

REFERENCES

1. Nauchnaia elektronnaia biblioteka eLibrary.ru. <https://www.elibrary.ru/>
2. Rossiiskii indeks nauchnogo tsitirovaniia.
https://www.elibrary.ru/project_risc.asp

3. *Ufimtseva A.* Natsionalnaia sistema otsenki nauki mozhet byt postroena na dannykh RINTs. Glava Elibrary.ru rasskazal, smozhet li Rossiia otkazatsia ot mezhdunarodnykh baz nauchnykh zhurnalov // *Gazeta.ru*. 14.03.2022.

<https://www.gazeta.ru/science/2022/03/14/14624287.shtml>

4. Reshenie arbitrazhnogo suda RF ot 22.02.2022 g. po delu № A40-267527/21-12-1844. URL: https://kad.arbitr.ru/Document/Pdf/9d4cf831-4442-4cb5-9519-b4686ab972e8/4a5e8937-4412-4657-a087-5a901de76d35/A40-267527-2021_20220222_Reshenija_i_postanovlenija.pdf

5. O proekte Russian Science Citation Index.

https://www.elibrary.ru/rsci_about.asp

6. DBLP computer science bibliography. <https://dblp.org/>

7. *Mazov N.A., Gureev V.N., Kalenov N.E.* Nekotorye otsenki spiska zhurnalov Russian Science Citation Index // *Vestnik RAN*. 2018. Tom 88, № 4. S. 322–332.

<https://doi.org/10.7868/S0869587318040047>

8. Akademicheskii reiting zhurnalov RSCI.

<http://www.ras.ru/news/shownews.aspx?id=6ddc26ee-044b-4cc9-99a0-96670f39ac13>

9. Predstavlenie Akademicheskogo reitinga na zasedanii Prezidiuma RAN 23 marta 2021 g. Vystuplenie A.R. Khokhlova.

<https://www.youtube.com/watch?v=LtFijzBOO0E> (4:12:45)

10. *Polilova T.A.* RAN sozdaet akademicheskii reiting nauchnykh zhurnalov. // *Nauchnyi servis v seti Internet: trudy XXII Vserossiiskoi nauchnoi konferentsii (20-23 sentiabria 2021 g., onlain)*. M.: IPM im. M.V. Keldysha, 2021. S. 279–296.

<https://doi.org/10.20948/abrau-2021-32>;

<https://keldysh.ru/abrau/2021/theses/32.pdf>

11. *Polilova T.A.* Reitingi zhurnalov v RINTs kak instrumenty analiza i vliianiia // *Preprinty IPM im. M.V. Keldysha*. 2021. № 40. 35 s.

<https://library.keldysh.ru/preprint.asp?id=2021-40>;

<https://doi.org/10.20948/prepr-2021-40>

12. *Polilova T.A., Gorbunov-Posadov M.M.* Reiting zhurnala v bibliograficheskoi baze // *Elektronnye biblioteki*. 2021. T. 246 №6. S. 1060–1089.

<https://doi.org/10.26907/1562-5419-2021-24-6-1060-1089>

13. *Gorbunov-Posadov M.V., Polilova T.A.* The Problem of Selecting a Reference Group for the Thematic Rating of Scientific Journals // CEUR Workshop Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). 2022. Vol. 3066. P. 90–100. <http://ceur-ws.org/Vol-3066/paper9.pdf>

14. *Polilova T.A.* Preprinty IPM im. M.V. Keldysha v diagrammakh Science Space // Preprinty IPM im. M.V. Keldysha. 2022. № 27. 38 s. <https://doi.org/10.20948/prepr-2022-27>; <https://library.keldysh.ru/preprint.asp?id=2022-27>

15. Herfindahl–Hirschman Index. From Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Herfindahl%E2%80%93Hirschman_Index

16. *Potter I.* Introducing Citation Topics in InCites – Clarivate. Dec 3, 2020. URL: <https://clarivate.com/blog/introducing-citation-topics/>

17. *Kasianov P.* Klassifikator predmetnykh oblastei Citation Topics v sisteme In-Cites. <https://youtu.be/58g6dQyJCW8>

18. Razdel “Novaia sistema akademicheskikh nadbavok” na saite NIU VShE. <https://www.hse.ru/science/scifund/an2021/>

19. Naukometricheskii tsentr NIU VShE. <https://scientometrics.hse.ru/>

20. Utverzhdennyy “Belyi spisok” nauchnykh zhurnalov. 2022. <https://new.ras.ru/activities/news/utverzhdennyy-belyy-spisok-nauchnykh-zhurnalov/>

21. *Beliaeva S.* Togo i zhdali: “Belyi spisok” nauchnykh zhurnalov nikogo ne udivil // Poisk. №47 (1745) ot 18 noiabria 2022. S. 4. https://poisknews.ru/wp-content/uploads/2022/11/poisk_47_20221118.pdf

22. “Belyi spisok” RAN na saite Rossiiskogo tsentra nauchnoi informatsii. <https://journalrank.rcsi.science/ru/record-sources/>

23. Vlasti vveli moratorii na uchet nauchnykh statei v mezhdunarodnykh izdaniyakh // RBK. 21 marta 2022 g. <https://www.rbc.ru/society/21/03/2022/6238915c9a794774a68b0245>

24. *Torkunov A.* Obshchestvennyye nauki v Rossii: zhurnaly i indeksy. 2022. <https://russiancouncil.ru/analytics-and-comments/comments/obshchestvennyye-nauki-v-rossii-zhurnaly-i-indeksy/>

25. *Kurilova A., Mishina V.* VAK gotovit reformu trebovaniy k nauchnym publikatsiyam soiskatelei stepenei // Vedomosti. 29 noiabria 2022.

<https://www.vedomosti.ru/society/articles/2022/07/25/932990-vak-reformu-trebovaniy-publikatsiyam>

26. VAK podelil nauchnye zhurnaly na tri kategorii. 16.12.22.

<https://poisknews.ru/science-politic/vak-podelil-nauchnye-zhurnaly-na-tri-kategorii/>

СВЕДЕНИЯ ОБ АВТОРЕ



ПОЛИЛОВА Татьяна Алексеевна – старший научный сотрудник Института прикладной математики им. М.В. Келдыша РАН, доктор физико-математических наук, лауреат Премии Президента РФ в области образования;

Tatyana Alekseevna POLILOVA – senior researcher of the Keldysh Institute of Applied Mathematics Russian Academy of Sciences.

email: polilova@keldysh.ru.

ORCID: 0000-0003-4628-3205

Материал поступил в редакцию 15 декабря 2022 года

УДК 330.8

ЭТАПЫ ТРУДНОГО ПУТИ

(о компьютеризации экономических исследований)

Ю. Е. Поляк^[0000-0001-8411-335X]

Федеральное государственное бюджетное учреждение науки Центральный экономико-математический институт Российской академии наук, 117418 Москва, Нахимовский пр. д.47

polak@cemi.rssi.ru

Аннотация

Работа посвящена переходу отечественной экономической науки от марксистских догм советского периода к широкому применению экономико-математического моделирования и компьютерной техники. Подчёркнута роль в этом процессе Центрального экономико-математического института (ЦЭМИ). Основанный в 1963 году, он стал лидером того периода среди институтов отделения экономики Академии наук. Обсуждены некоторые работы ЦЭМИ в области инфраструктурного и информационного обеспечения экономических исследований. Приведены сведения о ведущих экономистах прошлых лет.

Ключевые слова: *история отечественной экономики, ЦЭМИ, компьютерные сети, каталоги сетевых ресурсов, Соционет*

Первые попытки применения компьютерной техники для экономических расчетов и исследований относятся к концу 1950-х годов. В 1958 году А.И. Китов писал: «Электронные цифровые машины получают всё большее применение для экономических расчетов в области планирования, учета, статистики и для механизации административно-управленческой работы ... Вычислительные центры должны быть связаны в единую систему автоматической информационной и вычислительной службы, которая будет обеспечивать нужды всех учреждений и организаций в необходимой научной, технической, экономической и другой информации и выполнение вычислительных работ» [1]. В том же году была создана первая в стране лаборатория АН СССР по применению математических методов в экономических исследованиях и планировании (академик В.С. Немчинов).



21 мая 1963 года вышло постановление ЦК КПСС и СМ СССР № 564 «Об улучшении руководства внедрением вычислительной техники и автоматизированных систем управления в народное хозяйство», а 19 июля 1963 г. Постановлением президиума АН СССР № 519 на базе лаборатории Немчинова был образован Центральный экономико-математический институт АН СССР (ныне ЦЭМИ РАН), его директором стал академик Н.П. Федоренко. Постановлением № 603 от 11 октября 1963 г. была утверждена структура института. В качестве главной цели при создании института были провозглашены внедрение математических методов и ЭВМ в практику управления и планирования, создание теории оптимального управления народным хозяйством¹. К этому времени и в мире, и в стране происходили события, обозначившие возникновение цифровой цивилизации, в которой мы живем сегодня. ЭВМ, развивавшиеся до этого лишь для средств вооружения, направились во все сферы жизни. В этот период появились Институт экономики и организации промышленного производства СО АН СССР, Институт кибернетики АН УССР. Экономико-математическая тематика и применение ЭВМ в народном хозяйстве заняли важное место в работе ВЦ АН СССР, Института электронных управляющих машин; началась подготовка кадров экономистов-математиков и специалистов по экономической кибернетике в МГУ, НГУ, МИНХ им. Плеханова.

В ЦЭМИ была создана одна из лучших в стране организаций, занимающихся

¹ <http://www.cemi.rssi.ru/about/how>

непосредственно разработкой математических методов и прикладного программного обеспечения для решения задач управления экономикой [2]. Институт называли флагманом отрасли, лидером в развитии экономико-математического моделирования.

Однако исследования ЦЭМИ постоянно подвергались нападкам со стороны консервативно и догматически настроенных работников аппарата ЦК КПСС, органов экономического управления, секции общественных наук президиума АН СССР, не желавших отказываться от устаревающих методов хозяйствования. Внедрение в практику новых подходов вызывало крайнее раздражение этих ретроградных кругов [3]. Большой удар по экономико-математическому направлению был нанесён в июне 1983 года на пленуме ЦК КПСС, когда секретарь ЦК К. Черненко «обрушился с явно предвзятой критикой на ЦЭМИ» [4, с. 408]. Буквально было сказано следующее: «Многого мы ожидали от созданных еще в 60-х годах Института социологических исследований и Центрального экономико-математического института Академии наук СССР. Но до сих пор так и не дождались обстоятельных конкретных исследований социальных явлений и актуальных экономических проблем. В деятельности этих институтов наглядно проявились недостатки, в той или иной степени свойственные и некоторым другим научным учреждениям: замыкание в собственных «диссертационных» и групповых интересах, мелкотемье, слабость в них партийного влияния. Отделу науки и учебных заведений ЦК, президиуму и секции общественных наук АН СССР, партийным организациям следует уделять большее внимание вопросам планирования и организации исследований, стиля и методов работы научных коллективов»². Казалось бы, ничего особенного: далёкий от науки человек высказал своё (или чьё-то ещё) мнение, не обязательно верное. Но тогда по заведённому ритуалу тут же включились «верные ленинцы» из руководства Академии наук. Начались заседания президиума, секции общественных наук, сессии совета по координации научной деятельности; заработали комиссии по проверке работы института. В итоге были выработаны «мероприятия по коренному улучшению» работы – так, было предложено «осуществить реорганизацию Центрального экономико-математического института, который стал слишком громоздким и плохо управляемым, выделив из него

² <https://www.gensek.ru/publications/speeches/420.html>

подразделения, занимающиеся перспективами социально-экономического развития в связи с научно-технической революцией»³. В результате этой реорганизации из ЦЭМИ выделился Институт экономики и прогнозирования научно-технического прогресса (в настоящее время – Институт народнохозяйственного прогнозирования), а позднее – Институт социально-экономических проблем народонаселения и Институт проблем рынка. Из института ушёл ряд ведущих учёных, а академик Н.П. Федоренко был снят с поста директора.

Значительно трагичнее сложилась судьба учёных, занимавшихся экономико-математическими исследованиями в первые годы Советской власти. У них были заметные достижения – так, был разработан межотраслевой баланс производства и распределения продукции в народном хозяйстве страны за 1923/24 хозяйственный год; в докладе Г.А. Фельдмана «К теории темпов народного дохода» были предложены модели анализа и планирования развития экономики, заложены основы теории экономического роста. А.А. Богданов (Малиновский) ещё в 1897 г. опубликовал «Краткий курс экономической науки». В его работе «Тектология – всеобщая организационная наука» заложены основы системного подхода и теории организации как самостоятельной научной дисциплины. Тектология Богданова предвосхитила кибернетику Н. Винера и У. Эшби, общую теорию систем Л. фон Берталанфи и синергетику И. Пригожина. Его идеи, в свое время подвергшиеся гонениям, возродились и получили дальнейшее развитие в мировой науке. Они не потеряли своей актуальности, выступая фактически как теоретическая основа нынешней концепции устойчивого развития⁴. Н.Д. Кондратьев, автор книги «Большие циклы конъюнктуры» (1925), разработал теорию долговременных экономических циклов, нашедшую мировое признание. По Кондратьеву, циклы экономического развития (К-волны) повторяются каждые 45–60 лет в виде подъёмов и спадов в мировой экономике. Фаза роста сопровождается повышением уровня инфляции, ростом цен и зарплат. Фаза роста также сопровождается обострением борьбы за рынки и сырьё, социальными потрясениями (революциями, переворотами, войнами), резкими изменениями в социальной, политической, культурной, научной жизни людей.

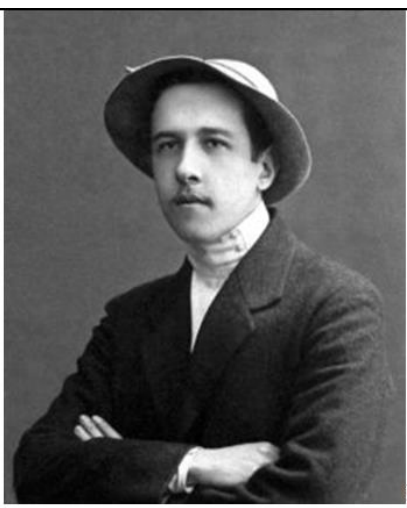
³ <https://ras.ru/FStorage/download.aspx?id=fa925b48-3c60-4219-a7e6-e98ff97d1762>

⁴ https://www.elibrary.ru/download/elibrary_9227776_36739022.pdf

Однако в 1930-е годы такие исследования в СССР были практически свёрнуты, а в результате разгромного процесса по «делу Трудовой крестьянской партии» (ТКП) несколько ведущих учёных-экономистов были расстреляны (Н.Д. Кондратьев, Л.Н. Литошенко, А.В. Тейтель, А.В. Чаянов и Л.Н. Юровский), другие приговорены к длительным срокам заключения.



Кондратьев Николай Дмитриевич (1892-1938) ☞



Чаянов Александр Васильевич (1888-1937) ☞



Юровский Леонид Наумович (1884-1938) ☞

В этой компании вполне мог оказаться и В.В. Леонтьев – в будущем знаменитый американский экономист, автор теории межотраслевого анализа, иностранный член Академии наук СССР, нобелевский лауреат 1973 года «за развитие метода „затраты — выпуск“ и за его применение к важным экономическим проблемам». Но он очень своевременно уехал из страны на лечение, его основные работы опубликованы в Германии и США; в СССР он успел только в возрасте 19 лет окончить Ленинградский университет и напечатать в журнале «Плановое хозяйство» статью «Баланс народного хозяйства СССР. Методологический разбор работы ЦСУ»⁵. В годы «перестройки» он неоднократно выступал в СССР с лекциями и сетовал, что власти пренебрегли его предложением помощи в проведении преобразований, переводу плановой экономики на рыночные рельсы. «Я туда больше не поеду. Они ничего не слушают».

⁵ https://istmat.org/files/uploads/28606/planovoe_hozyaystvo_1925_-12.pdf

В Российской империи родился ещё один нобелевский лауреат по экономике (1971), С.А. Кузнец. Согласно Википедии, он «оказал большое влияние на развитие целого ряда отраслей экономической мысли. С его именем связывают становление современной экономики как эмпирической научной дисциплины, развитие статистических методов исследований и появление количественной экономической истории». Образование Кузнец получил в Харькове, и в 1922 г. прибыл в США с готовой магистерской диссертацией, написанной на втором курсе. В Америке Кузнец следил за шедшей в СССР экономической дискуссией, много писал о советской статистике, издал сборник по проблемам статистических сопоставлений; в 1920-х годах рецензировал и переводил малоизвестные на Западе работы Н.Д. Кондратьева, А.Л. Вайнштейна⁶.



Харьковчане – нобелевские лауреаты

Советские экономисты вели борьбу за идеологическую чистоту в своих рядах и в довоенный период, и после войны. После «процесса ТКП» развернулась политическая травля «неправильных» экономистов, в которой приняли активное участие их коллеги. Формировались новые мораль и этика, ставшие типичными для всей советской науки. Новая советская (точнее, сталинская) экономическая

⁶ <https://web.archive.org/web/20131111215746/http://cdn.scipeople.com/materials/32916/Kuznets-i-kharkovskaya-shkola.pdf>

наука нарождалась именно в такой обстановке идеологического террора, предательства, классовой кровожадности и беспощадности. Советские партийные экономисты с верноподданническим усердием, ставшим со временем их неотъемлемой чертой, создавали эту «науку». Многие из них охотно приспособивались не только к конкретным и порой противоречивым положениям марксизма-ленинизма, но и ко всем изменениям политической конъюнктуры внутри страны, к изменяющимся формулировкам в решениях партийных съездов и пленумов. «Научное» обоснование получали не только бесчисленные «преимущества социализма», но и явные ошибки руководства страны⁷.

Перефразируя классика, заметим, что многие ошибки хуже, чем просто ошибки, это преступления. 26 мая 1947 года в СССР была отменена смертная казнь: «страна, победившая фашизм и занятая построением самого справедливого общества на земле, не должна была ассоциироваться с жестокостью»⁸. Но через три года подоспело «ленинградское дело», и «по просьбам трудящихся» расстрелы возобновились. Среди 26 погибших партийных и государственных руководителей привлекают внимание известные экономисты братья Вознесенские – Александр (ректор Ленинградского университета, министр просвещения РСФСР) и Николай (председатель Госплана, член Политбюро и заместитель председателя Совета министров СССР). Они обвинялись в измене Родине, участии в контрреволюционной организации и антисоветской агитации, хозяйственных преступлениях. Но, по мнению «примкнувшего» Д.Т. Шепилова, важными причинами стали борьба за власть в руководстве страны и доносы П. Федосеева (впоследствии вице-президента АН СССР, члена ЦК, обвинявшего ЦЭМИ во всех грехах в 1983 г. на заседании Президиума академии)⁹.

⁷ https://ecsocman.hse.ru/data/423/676/1219/glava_5.pdf

⁸ <https://life.ru/p/1399040>

⁹ <https://biography.wikireading.ru/179398>

	
Александр Алексеевич Вознесенский (1898-1950)	Николай Алексеевич Вознесенский (1903-1950)

Процессы конца 1930-х и конца 1940-х годов обошли стороной ещё одного известного учёного. Их избежал А.Л. Вайнштейн, уже находившийся в заключении, а всего в сталинских тюрьмах, лагерях и ссылках он провёл 18 лет. Его первые научные работы были посвящены вопросам экономики и математической статистики сельского хозяйства. В годы НЭПа вышли его первые монографии. Вайнштейн стал одним из первых отечественных интерпретаторов метода прогнозирования рыночной конъюнктуры, получившего название «экономического барометра». Он занимался применением методов гармонического анализа (рядов Фурье) к исследованию периодичности и прогнозированию колебательных процессов в экономике, а также анализом оперативной и долгосрочной народнохозяйственной динамики. Вайнштейн несколько раз встречался с Лениным, который на каком-то приёме якобы сказал: «Нам нужны такие умные молодые специалисты. Вам бы стать членом большевистской партии». Однако учёный отказался: «Я не верю в диктатуру пролетариата»¹⁰. После реабилитации А.Л. Вайнштейн стал одним из ведущих теоретиков в только что организованном ЦЭМИ АН СССР.

Ожесточённым нападкам подвергался Л.В. Канторович, будущий лауреат Нобелевской премии по экономике 1975 года (точнее – премии Шведского национального банка по экономическим наукам памяти Альфреда Нобеля, Sveriges Riksbanks pris i ekonomisk vetenskap till Alfred Nobels minne) за вклад в теорию оптимального распределения ресурсов. А в 1939 году, опубликовав свой метод ли-

¹⁰ <https://litbook.ru/article/654>

нейного программирования, он столкнулся с сопротивлением применению математических методов в экономике. Тогда считалось, что западная математическая школа в экономике – это антимарксистская школа, и математика в экономике – средство апологетики капитализма. Партийная номенклатура мстила Канторовичу за его острую и бескомпромиссную критику их беспросветного невежества¹¹.

Критики и ретрограды получили неожиданную поддержку со стороны руководителей предприятий, которым успешное применение революционных методов Канторовича часто не приносило счастья. При советской системе планирования «от достигнутого уровня» те, кто добивался значительной экономии сырья и материалов, получали на следующий год совершенно неподъёмные задания.

Открытию Канторовича противостояли не только объективные сложности использования метода, но и догматическая косность большинства советских экономистов, привыкших управлять экономикой при помощи лозунгов из Маркса. В новой сложной теории, смысл которой они не могли понять, идеологи увидели угрозу своему материальному благополучию. Применение науки наталкивалось на волюнтаризм начальствующих дураков, их невежество в науке¹². Канторовича подвергали принудительному лечению в психбольнице, травили в партийной печати. Так, выдвижение его книги «Экономический расчет наилучшего использования ресурсов» на соискание Ленинской премии названо в газете «Правда» совершенно неправильным¹³. Авторы статьи, «генералы от науки», обвиняют книгу Канторовича в том, что в ней «в основу ценообразования и планового руководства всей экономики положены своеобразно исчисляемые «объективно обусловленные оценки». Известно, что концепция этих «оценок» единодушно отвергнута в советской печати как теоретически ошибочная и практически бесплодная ... Теоретическая ошибочность всей концепции оценок Л.В. Канторовича состоит в том, что она явно противоречит трудовой теории стоимости». Привлекая в качестве аргументов идеологические лозунги, программу КПСС, курс партии на химиза-

¹¹ <https://proza.ru/2018/11/24/175>

¹² https://chronology.org.ru/Канторович,_Леонид_Витальевич

¹³ Струмилин С., Островитянов К. и др. В плену теоретических ошибок // Правда, 10 апреля 1964 года

цию, «механизацию и автоматизацию в сочетании с повышением культурно-технического уровня трудящихся», они завершают статью ударным абзацем, напоминающем донос:

«Не случайно буржуазная печать на Западе сразу же после выхода книги Л.В. Канторовича стала восхвалять его на все лады именно за выдвижение его оценок на место стоимости в марксистском понимании. Буржуазные экономисты тщательно пытаются доказать, что применение математики якобы несовместимо с марксистской экономической теорией, они хотели бы, чтобы математика осталась их монополией. В книге Л.В. Канторовича они видят подтверждение своих взглядов о том, что применение математики якобы требует отступления от марксизма, перехода на их теоретические позиции».

Появись такой текст лет на 25–30 раньше, последствия для Канторовича могли быть трагичными. Но в 1964 году, уже после этой статьи, он был избран академиком АН СССР, затем награждён двумя орденами Ленина и орденом Трудового Красного знамени¹⁴. А Ленинскую премию Канторович всё же получил, но годом позже.

		
<p><i>26-летний профессор Л.В. Канторович</i></p>	<p>К.С. Петров-Водкин. Портрет учёного Л.В. Канторовича. 1938</p>	<p>Л.В. Канторович, Москва, санаторий «Узкое». 1976. © А. Богданов</p>

¹⁴ https://ru.wikipedia.org/wiki/Канторович,_Леонид_Витальевич

Однако нападки и клевета на Л.В. Канторовича продолжались и после его смерти. Известный в прошлом литератор С. Залыгин в 1996 г. опубликовал в редактируемом им журнале автобиографические заметки, где, в частности, говорится: «Академик Канторович ... сказал мне, что он не может и не должен жить в стране, в которой он никому-никому не нужен, никем не понимаем, а в силу этого даже и презираем, и что при первой же возможности он покинет Советский Союз, поселится в Америке, по модели которой он разрабатывает систему математической экономики. Так и получилось: года через два Канторович эмигрировал в Америку, а это по тем временам был случай совершенно исключительный, ещё год-другой спустя он стал лауреатом Нобелевской премии»¹⁵. Вероятно, из-за подобных мемуаров появилось выражение «врёт, как очевидец». Во-первых, Л.В. Канторович не разрабатывал систему математической экономики по образцу Америки, а предложил в 1939 году принципиально новые математические методы организации и планирования производства, заложив основы линейного программирования. Во-вторых, он никогда не эмигрировал, в последние годы жил и работал в Москве, где и похоронен в апреле 1986 г. на Новодевичьем кладбище.

Так совпало, что весна 1986 года ознаменовала, по мнению автора, начало компьютеризации ЦЭМИ АН СССР и в целом отделения экономики (ОЭ) Академии наук. Именно тогда по указанию академика-секретаря ОЭ А.Г. Аганбегяна были выделены ресурсы (\$150000) для компьютерного перевооружения организаций ОЭ. Эта работа была поручена лаборатории персональных ЭВМ ЦЭМИ. Её сотрудники готовили спецификации, вели переговоры с поставщиками техники, добиваясь льготных условий поставки; выполняли приёмку оборудования, ввод его в эксплуатацию и распределение по институтам, а также обучение представителей институтов работе с новой техникой. В результате в институтах Отделения экономики появились персональные компьютеры IBM PC, а впоследствии IBM PC/XT и AT с процессорами Intel 80286 и 80386 соответственно; едва ли не первые в системе Академии наук лэптопы – портативные компьютеры фирм Toshiba и Bondwell с тактовой частотой 4.77 МГц и весом 8 кг, периферийное оборудование

¹⁵ https://magazines.gorky.media/novyi_mi/1996/12/moya-demokratiya.html

(в том числе редкие для того времени лазерные принтеры HP); накопители, математические сопроцессоры, сетевое оборудование и т. д. В дальнейшем лаборатория выполняла функции координационного и методического центра по использованию ПЭВМ, созданию фонда программного обеспечения. На лабораторном семинаре «Программные средства для экономических исследований» специалисты делали обзорные доклады о новых разработках в этой области. Был организован телекоммуникационный доступ к локальным и удалённым базам данных, велась работа по дополнительному оснащению института новым оборудованием и программными средствами.

В 1986 г. распоряжением Президиума Академии наук на институт были возложены функции компьютерной подготовки аспирантов и соискателей из секции общественных наук АН СССР – разработка программ и учебных материалов, проведение занятий и экзаменов по программе кандидатского минимума. Организация этой работы была поручена той же лаборатории персональных ЭВМ. На базе поставленной техники был создан класс, оснащённый 30-ю IBM-подобными компьютерами, объединёнными в сеть. Обучение по 100-часовой программе курса «Информатика и вычислительная техника» прошли более 70 групп из институтов отделений экономики и мировой экономики. Занятия в институтском компьютерном классе проводили сотрудники ЦЭМИ А.Б. Аранович, Т.С. Онучак, Ю.Е. Поляк, Н.А. Трофимова. В 1990 г. свыше 200 сотрудников аппарата Президиума АН СССР повысили свою квалификацию в сфере компьютерной грамотности по специально разработанной программе (эта работа велась сотрудниками лаборатории на технической базе научно-учебного комплекса АН). Проводилось обучение сотрудников института (среди слушателей первых групп были К.А. Багриновский, Е.Г. Ясин, Е.Г. Гольштейн, С.А. Алексащенко и др.), разрабатывались методические материалы для обучения. В 1988 г. в компьютерном классе ЦЭМИ занимались сотрудники Института философии (которые через много лет благодарили автора за первые навыки общения с компьютером, помогавшие им выжить, когда основная профессия не кормила). За организационно-методическим содействием обращался руководитель ВЦСПС Г.И. Янаев. В 1989 г. на базе лаборатории был организован филиал кафедры социальной информатики МГУ.

Только с появлением персональной вычислительной техники и интернета

моментам возник массовый интерес к их возможностям. Вспоминается, как пожилые учёные, воспитанные на марксистско-ленинских догмах и далёкие от компьютеров, просили помочь им в переписке с детьми, обучающимися за рубежом, и отправке своих сочинений на запад для публикации. В ЦЭМИ в 1990-е годы началось активное использование персональных компьютеров не только для печати текстов, но и для решения управленческих задач.

Когда через несколько лет персональные компьютеры вошли в повседневную практику большинства академических организаций, существование такого специализированного подразделения как лаборатория ПЭВМ стало неактуальным. Однако появилась другая проблема. В отсутствие средств навигации (поисковые системы Апорт и Рамблер появились только в 1996 г., Яндекс – в 1997, Google – в 1998) возникла острая необходимость в каталогах и справочниках (как печатных, так и электронных) с рекомендациями о полезных информационных ресурсах (состояние рунета того периода подробно описано в [16, 20]). Сотрудник лаборатории сетевых информационных ресурсов А.Н. Дыбенко разработал на MS Access оболочку базы данных «Интернет в России, Россия в интернете». Информационным наполнением и редактированием записей в базе занимались сотрудники ЦЭМИ, а также (в 1996–2002 гг.) студенты вечернего отделения МГУ, которым Ю.Е. Поляк читал лекции по информатике, – всего более 30 человек. К этому времени количество записей превысило 50 тысяч. База зарегистрирована в Государственном регистре баз данных (№ госрегистрации 0229905944); на её основе создан ряд уникальных онлайн-информационных проектов, а также печатных справочников [5–10] и каталогов ресурсов. Среди каталогов выделяется «АУ!» [11, 12, 21] – лучший и наиболее полный для своего времени. Он был представлен на семинаре РОЦИТ в ноябре 1996 г. и в отсутствие развитых средств сетевой навигации стремительно сделался популярным. Этому способствовали особенности «АУ!», отмеченные в 1997 году руководителями РОЦИТ А.Ю. Зотовым и В.Э. Соркиным: это «полнота и качество представленных в нем описаний. Качество описаний определяется сформированной и выверенной методикой работы аналитической группы, тщательно отработанными рубрикаторами, сочетающими полноту и сбалансированность. По мнению РОЦИТа, качество описаний делает



каталог «АУ!» уникальным даже в сравнении с таким известным и близким аналогом, как поисковая система Yahoo!»¹⁶.

«АУ!» – не единственный информационный продукт, созданный на основе базы данных «Интернет в России, Россия в интернете». В 1999 г. к нам обратились руководители компании «Яндекс», с которыми у нашего коллектива были давние дружеские отношения. Они решили для повышения релевантности при поиске в интернете создать каталог ресурсов совершенно нового типа.

На рубеже веков стало окончательно ясно, что появившиеся к тому времени поисковые машины нового поколения с их миллионами индексируемых документов и постоянной актуализацией баз данных выиграли заочное соревнование в скорости и объёмах с живыми редакторами. Из-за колоссального роста объёмов сетевой информации стало невозможно продолжать актуализацию универсального каталога на прежнем высоком уровне. Началась конвергенция каталогов и поисковых систем, позволившая им приобрести лучшие черты друг друга [13]. Ограничение области поиска определенной тематической категорией экономит время поиска и в то же время повышает релевантность его результатов. Тщательный ручной мониторинг встроенного каталога в сочетании с быстрым автоматическим обходом сайтов, выполняемым поисковыми роботами, обеспечивает высокую степень соответствия запроса и результата.

Работа началась в конце 1999 года. Наш коллектив совместно со специалистами «Яндекса» разработал концепцию каталога и принципы классификации, выполнил первоначальное наполнение базы данных и обучение редакторов. В основу был положен ряд новых решений; одно из основных – применение многомерной рубрикации. Сходные принципы положены в основу фасетной классификации [14, 15], но в интернете они ранее не использовались. Сайты в каталоге были отсортированы по индексу цитирования – числу упоминаний в интернете.

¹⁶ Из предисловия к печатному справочнику интернет-ресурсов, подготовленному для издательства «Открытые системы» в 1998 году (выпуск не состоялся из-за известных событий на финансовом рынке)

	
<p>Труды лаборатории сетевых информационных ресурсов (1999 г.)</p>	<p>Генеральный директор «Яндекса» А.Ю. Волож и студентки МГУ, лаборанты ЦЭМИ РАН В. Мацкевич и И. Хренова 27.06.2000. Фото автора</p>

Новый каталог был интегрирован с поиском Яндекса. К лету 2000 г. в каталоге Яндекса было уже более 40000 российских интернет-ресурсов, что соответствует примерно 300 тысячам единиц описания в обычном каталоге [16]. Лаборатория сетевых информационных ресурсов ЦЭМИ работала над концепцией и пополнением каталога Яндекса несколько лет. После ряда модернизаций каталог продолжал действовать до 2018 года; к этому времени он содержал 108 тысяч адресов¹⁷.

Следует упомянуть и о работе над системой рейтингов Рамблер Top100¹⁸, для которой в 2008 году были приглашены специалисты ЦЭМИ. Были разработана новая структура рубрикатора, создана система словарей, проведён анализ поисковых запросов и посещаемости разделов. Но экономическая ситуация в стране (и в Рамблере) не позволила реализовать всё задуманное [21].

Задолго до широкого распространения интернета институт вёл работы по соединению с мировыми компьютерными сетями. 25–26 сентября 1990 года в Москве с участием сотрудников ЦЭМИ прошло первое всесоюзное совещание по

¹⁷ <http://wikireality.ru/wiki/Яндекс.Каталог>

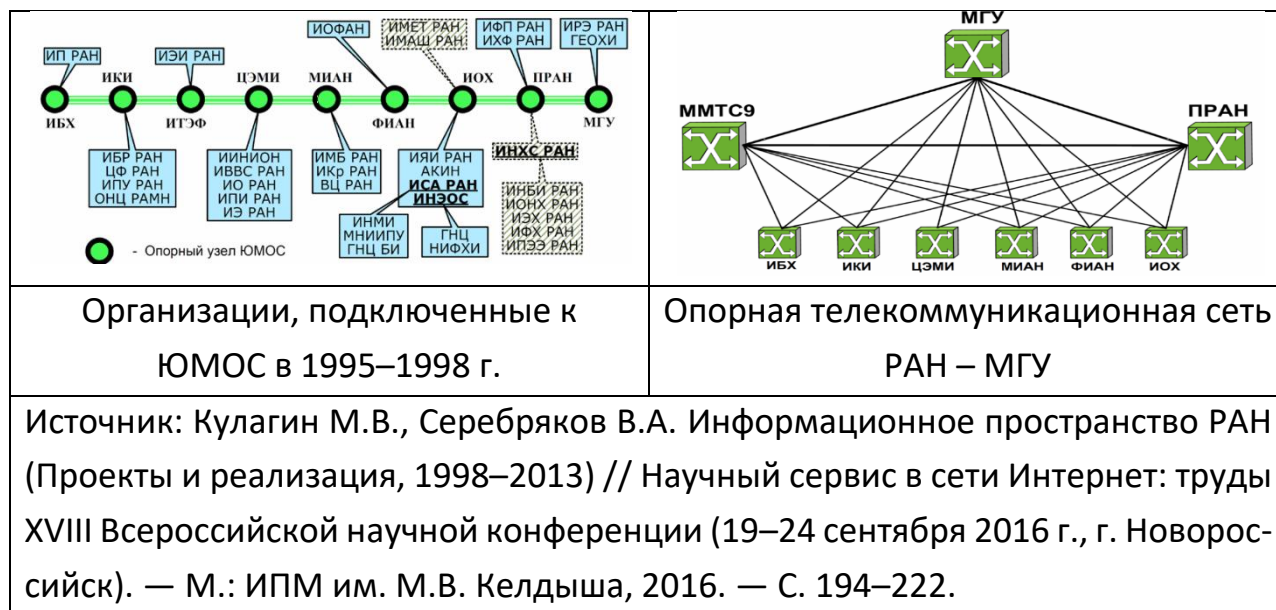
¹⁸ <https://top100.rambler.ru>

организации подключения академических организаций к международной компьютерной сети EARN/BITNet¹⁹ и созданию SUEARN — советской части EARN. В 1992 г. вступили в эксплуатацию её первые узлы: в ИОХ (SUEARN2) и ЦЭМИ (SUCEMI), что позволило подключиться к международному информационному обмену и другим институтам РАН.

С конца 1980-х годов научно-образовательные организации начали создавать центры доступа в интернет. В 1991 г. была высказана идея строительства волоконно-оптической линии связи, которая связала бы ведущие научно-образовательные учреждения, расположенные в основном вдоль линии метро «Ленинский проспект – Калужская» на юго-западе Москвы, где находятся МГУ, многие институты РАН и научные организации других ведомств. Эта первая телекоммуникационная магистраль, появившаяся в 1994 году и получившая название ЮМОС (Южная московская опорная сеть), стала объединяющим центром для научно-образовательных организаций. 6 января 1993 года прошло первое заседание группы по разработке сети ЮМОС, в которую входил и представитель ЦЭМИ РАН. Институту в ЮМОС отводилась важная роль, так как ЦЭМИ являлся опорным узлом и обеспечивал подключение к сети институтов отделения общественных наук РАН (ООН РАН) и некоторых других близлежащих научных организаций. В число первых вошли бывшие отделы ЦЭМИ, выделившиеся из него в 1985, 1988 и 1990 гг. и ставшие самостоятельными институтами — ИНП РАН, ИСЭПН РАН и ИПР РАН.

Весной 2000 года было принято решение о финансировании проекта новой опорной сети, получившей название «Опорная телекоммуникационная сеть РАН – МГУ» (ОТС РАН – МГУ). В развитии новой сети активное участие приняли и сотрудники ЦЭМИ.

¹⁹ EARN - European Academic & Research Network (позднее вошла в TERENA – Trans-European Research and Education Networking Association и затем в GEANT – GEometry ANd Tracking, geant.net); BITNet - Because It's Time Network - «потому, что пришло время»



Основной задачей проекта являлось создание высокоскоростной транспортной среды (Gigabit Ethernet) на современном сетевом оборудовании, обеспечивающем дальнейшее развитие сервисной инфраструктуры корпоративной сети РАН, являющейся базой для реализации инновационных технологий – Grid-проектов, ориентированных на распределённые вычисления: построение вычислительных кластеров и развитие параллельного программирования. Создание гигабитного узла дополнило возможности использования интернет-узла ЦЭМИ другими институтами РАН, присоединёнными к уже действовавшему в ЦЭМИ узлу академической сети ОТС РАН-МГУ и узлу опорной сети ЮМОС. Эта инфраструктура обеспечивает научным коллективам доступ к удалённым информационным и вычислительным ресурсам в нашей стране и за рубежом, организацию оперативного обмена научной информацией и создание на основе современных информационных технологий условий для проведения совместных научно-исследовательских работ.

В заключение опишем ещё одну разработку, сыгравшую важную роль в снабжении исследователей научной информацией и поддерживавшуюся в ЦЭМИ. Это информационная система Соционет²⁰ – агрегатор коллекций научных

²⁰ <https://socionet.ru>

материалов, а также инструмент для создания собственного рабочего пространства. Соционет основан на парадигме открытой науки и представляет собой автоматически пополняемую базу статей.

Несколько слов об открытой науке. Движение за открытый доступ к научным публикациям (Open Access) зародилось в 1990-е годы в качестве альтернативы всевластию научных издательств. Его начало было положено П. Гинспаргом в 1991 году, когда он основал открытый депозитарий arXiv²¹ на базе Лос-Аламосской национальной лаборатории. На данный момент, это крупнейший бесплатный архив научных публикаций в области физики, математики, астрономии, информатики и биологии. В феврале 2002 года на Будапештской конференции по открытому доступу²² была сформулирована суть понятия «открытый доступ»: бесплатный (free), немедленный (immediate), постоянный (permanent), полнотекстовый (fulltext), онлайн-доступ (online) к научным публикациям [21]. Логическим продолжением идеи открытого доступа стала открытая наука (Open Science) как реакция на высокие подписные цены научных журналов и следствие важных изменений в практике научной работы и в научных коммуникациях под влиянием новых информационных технологий. Цели открытой науки: прозрачность в методологии экспериментов, наблюдений и сбора данных; публичная доступность и повторное использование научных данных; доступность и прозрачность научных коммуникаций; использование онлайн-средств для поддержки научного сотрудничества²³.

Материалы в Соционет загружаются с таких ресурсов, как CyberLeninka.ru, редакции научных журналов и т. д. Отличительная особенность ресурса заключается в возможности непосредственного взаимодействия автора и читателя. Многие функции ресурса Соционет уникальны как для российского, так и для международного научного пространства²⁴. Система удовлетворяет всем требованиям «идеальной» глобальной системы научных коммуникаций, основанной на современных информационных технологиях и обеспечивающей циркуляцию в научном

²¹ <https://arxiv.org>

²² <https://www.budapestopenaccessinitiative.org>

²³ <https://keldysh.ru/abrau/2015/304.pdf>

²⁴ <https://aspirantura.hse.ru/soc/news/184733732.html>

сообществе создаваемых результатов исследований и их доведение до потребителя [17].

Система Соционет свободно доступна пользователям с 2000 г. Её разработка началась ещё в прошлом тысячелетии [18]. Руководитель рабочей группы – сотрудник Института экономики и организации промышленного производства Сибирского отделения Российской академии наук (ИЭОПП СО РАН)²⁵ Сергей Иванович Паринов (в настоящее время – главный научный сотрудник ЦЭМИ РАН). Соционет предоставляет свободный доступ к большому массиву научных публикаций по различным дисциплинам, среди которых наибольшую долю занимают публикации социально-экономического характера. Количество публикаций исчисляется миллионами. Зарегистрированные пользователи могут публиковать в системе свои материалы и информационные объекты, а также пользоваться её разнообразными бесплатными сервисами. Система предоставляет доступ к электронным объектам различного рода (монографии и статьи из периодики, научные отчёты, тексты и презентации докладов, авторефераты диссертаций и полные их тексты, рабочие записки и многое другое). Кроме того, доступны сведения о персонах – авторах публикаций, а также об организациях, с которыми ученые аффилированы.

Соционет формирует виртуальную научную среду в духе открытой науки для создания и использования информационных ресурсов и сервисов, адресованных профессиональным научным сообществам. Разработка финансировалась различными фондами и организациями. С 2000 по 2007 гг. это были гранты Фонда Форда и Европейской комиссии, затем – гранты и помощь от российских организаций. До 2006 г. хостинг для системы Соционет предоставлял ИЭОПП СО РАН, затем – ЦЭМИ РАН. Наиболее активными пользователями Соционета стали 29 институтов Отделения общественных наук РАН, для которых система Соционет в том числе играет роль шлюза для передачи данных в Единое научное информационное пространство РАН. Уже к февралю 2007 г. институты ООН РАН создали свои онлайн-репозитории и выложили в открытый доступ в рамках Соционета около 30 тысяч материалов, включая 41 профиль 32 организаций (подразделений), 466 профилей учёных. Практически все институты ведут ленты новостей, коллекции

²⁵ <https://www.ieie.su>

статей своих сотрудников, описания значимых опубликованных материалов (книги и монографии), авторефераты защищаемых диссертаций и т. п. [19]



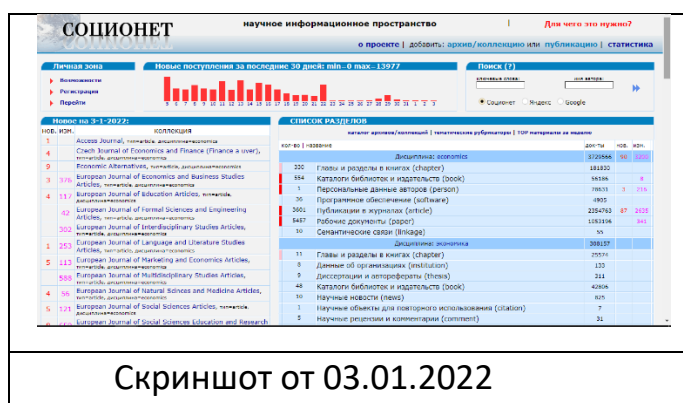
С.И. Паринов на конференции «Научный сервис в сети Интернет». Абрау-Дюрсо, 2006. Фото автора

К сожалению, на момент подготовки этого текста (январь 2023) в связи с решением Мосгорсуда²⁶ и требованием Роскомнадзора доступ к накопленным в Соционете данным о научных публикациях закрыт, система перенесена из компьютерной сети ЦЭМИ РАН на хостинг NIC.RU. Иск был подан неким ООО в составе двух человек²⁷, поставившим на поток тяжбы с организациями, занимающимися изданием и распространением научно-образовательной информации (за три года на сайтах Мосгорсуда и Арбитражного суда зарегистрировано порядка 100 исков с требованиями выплаты штрафов за нарушение авторских прав). При этом штраф за «нарушения» предъявлен С.И. Паринову, а также Российской академии наук и ЦЭМИ РАН. Мосгорсуд решил, что размещение в системе Соционет библиографической информации нарушает некие авторские права. Однако в Соционете представлены только метаданные публикаций; самих текстов там нет и не было. С таким же успехом можно утверждать, что информация на стандартных карточках в

²⁶ <https://mos-gorsud.ru/mgs/defend/get/document/36a2f1c0-b193-11ec-a0ef-33fdb7278959>

²⁷ https://www.audit-it.ru/contragent/1137746416390_ooo-mediamuzyka

библиотечном каталоге тоже нарушает чьи-то авторские права. Чтобы обезопасить себя от возможных новых исков и не создавать прецедент для других подобных «серийных истцов», команда Соционет 6 апреля 2022 г. полностью закрыла доступ к системе. Российские учёные лишились (надеемся, временно) крупной научной информационной системы с ежедневной аудиторией порядка 2 тысяч человек, которая предоставляла данные о 4 миллионах публикаций. В 2023 г. предостоят новые судебные разбирательства.



Компьютеризация экономики давно стала свершившимся фактом. История этого процесса продолжает вызывать интерес. Мы попытались показать, как на протяжении жизни одного поколения экономическая наука превратилась из набора политэкономических заклинаний с обочины мировой экономической мысли в современную научную дисциплину с теоретической базой, авторитетными специалистами, технической и программной поддержкой. ЦЭМИ продолжает активно участвовать в этом процессе, несмотря на объективные и искусственные препятствия.

СПИСОК ЛИТЕРАТУРЫ

1. *Китов А.И.* Электронные вычислительные машины. Серия VIII. Вып. II. №23. М.: Знание. 1958. 31 с.
URL: http://it-history.ru/images/b/b0/Kitov_A_Elektroonye_vichislitelnie_mashiny.1958.pdf
2. *Ким К.В.* Разработки ЦЭМИ в области компьютеризации и автоматизи-

рованных систем управления экономикой (1963–1985 гг.) // Экономика и математические методы. 2018. №3. С. 159–173.

3. *Афанасьев А., Баранов Э.* К 100-летию со дня рождения академика Н.П. Федоренко // Экономика и математические методы. 2018. №3. С. 4–12

4. *Лопатников Л.И.* Экономико-математический словарь: Словарь современной экономической науки. М.: Дело, 2003. 520 с.

5. Мировая сеть Internet: применение в науке и бизнесе (под ред. Ю.М. Горностаева и Ю.Е. Поляка) // Технологии электронных коммуникаций, Т. 59. М.: Эко-Трендз, 1994.

6. *Вовченко Т.О., Гунькин А.А., Поляк Ю.Е.* Компьютерные сети и их информационные ресурсы. Препринт ЦЭМИ РАН. М., 1995. 72 с.

7. Информационные ресурсы интернета (краткий справочник) // Технологии электронных коммуникаций, Т. 68. М.: Эко-Трендз, 1996. 60 с.

8. *Поляк Ю.Е., Сигалов А.В.* Жёлтые страницы Internet'98. Русские ресурсы. СПб: «Питер», 1998. 600 с.

9. Russian Internet Directory. 2nd ed. (Editors: Yuri Polak, Yuri Gornostaev). ICSTI, Moscow, 1998.

10. *Поляк Ю.Е.* Информационные ресурсы российского сегмента интернета // Информатизация и связь. 1998. № 1. С. 37–46.

11. *Куликов В.В.* Как мы делали «АУ!» // Компьютерра. №28 (257), 1998.

12. *Куликов В.В., Поляк Ю.Е.* Каталог русскоязычных ресурсов интернета «АУ!» // Телематика'98. СПб.: 1998. С. 334–335.

13. *Polak Y.* Information search in Russian Internet on the eve of 2000 // Internet: Technologies and Services. The 1st IEEE International Conference and Exhibition, 1999.

14. *Ранганатан Ш.Р.* Классификация двоеточием. Основная классификация. Пер. с англ. под. ред. Т.С. Гомолицкой и др. М.: ГПНТБ СССР, 1970.

15. *Foskett A.C.* The future of faceted classification // The future of classification, ed. by R. Marcella and A. Maltby, Aldershot: Gower, 2000. P. 69–80.

16. *Парахина О.В.* Каталоги интернет-ресурсов и их место среди поисковых инструментов // Материалы конференции представителей региональных научно-образовательных сетей «RELARN-2001». СПб, 2001. С. 93–96.

17. Когаловский М.Р., Неволин И.В., Паринов С.И. Развитие научных коммуникаций как условие модернизации оценки научной результативности // Эпистемология и философия науки. 2017. Т. 51. № 1. С. 188–205.

18. Паринов С., Ляпунов В., Пузырёв Р. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов // Электронные библиотеки. 2003. Т. 6. № 1. URL: <https://elbib.ru/article/view/222>

19. Паринов С.И. Онлайн-будущее науки: наукометрическая сигнальная система: Препринт WP2/2007/01. М.: Издательский дом ГУ ВШЭ, 2007. 52 с.

20. Поляк Ю.Е., Сигалов А.В. Интернет в России: сетевая инфраструктура и информационные ресурсы // Библиотечные компьютерные сети: Россия и Запад. Современные тенденции корпоративной работы библиотек в сетях передачи данных. Сборник статей. Москва, 1998. С. 10–34.

21. Акиншин А.А., Поляк Ю.Е. Вклад ЦЭМИ в становление рунета. Информационные проекты // Вестник ЦЭМИ РАН. 2022. Т. 5, Вып. 2. URL: <https://cemi.jes.su/s265838870019963-2-1>

STAGES OF THE DIFFICULT WAY (on the computerization of economic research)

Y.E. Polak^[0000-0001-8411-335X]

Central Economics and Mathematics Institute of the Russian Academy of Sciences, 47 Nakhimovski Pr. Moscow 117418 Russia

polak@cemi.rssi.ru

Abstract

The paper is devoted to the transition of domestic economic science from the Marxist dogmas of the Soviet period to the widespread use of economic and mathematical modeling and computer technologies. The role of the Central Economics and Mathematics Institute in this process is emphasized. Founded in 1963, it became the leader of that period among the institutes of the Economics Department of the Academy of Sciences. Some works of CEMI in the field of infrastructure and information

support of economic research are discussed. Information about some leading economists of past years is given.

Keywords: *history of the domestic economy, CEMI, computer networks, directories of network resources, Socionet*

REFERENCES

1. *Kitov A.I.* Elektronnye vychislitel'nye mashiny. Seriya VIII. Vyp. II. №23. M.: Znanie. 1958. 31 s. URL: http://it-history.ru/images/b/b0/Kitov_A_Elektronnye_vychislitelnie_mashiny.1958.pdf
2. *Kim K.V.* Razrabotki CEMI v oblasti komp'yuterizacii i avtomatizirovannyh sistem upravleniya ekonomikoj (1963–1985 gg.) // *Ekonomika i matematicheskie metody.* 2018 №3. С. 159–173.
3. *Afanas'ev A., Baranov E.* K 100-letiyu so dnya rozhdeniya akademika N.P. Fedorenko // *Ekonomika i matematicheskie metody.* 2018. №3. S. 4–12
4. *Lopatnikov L.I.* Ekonomiko-matematicheskij slovar': Slovar' sovremennoj ekonomicheskoy nauki. M.: Delo, 2003. 520 s.
5. *Mirovaya set' Internet: primenenie v nauke i biznese (pod red. Y.M. Gornostaeva i Y.E. Polaka)* // «Tekhnologii elektronnyh kommunikacij», T. 59. M.: Eko-Trendz, 1994.
6. *Vovchenko T.O., Gun'kin A.A., Poyak Y.E.* Komp'yuternye seti i ih informacionnye resursy. Preprint CEMI RAN. M., 1995. 72 s.
7. *Informacionnye resursy interneta (kratkij spravochnik)* // *Tekhnologii elektronnyh kommunikacij*, T. 68. M.: Eko-Trendz, 1996.
8. *Polak Y.E., Sigalov A.V.* Zhyoltye stranicy Internet'98. Russkie resursy. SPb: «Piter», 1998. 600 s.
9. *Russian Internet Directory.* 2nd ed. (Editors: Yuri Polak, Yuri Gornostaev). IC-STI, Moscow, 1998.
10. *Polak Y.E.* Informacionnye resursy rossijskogo segmenta interneta // *Informatizaciya i svyaz'.* 1998. № 1. S. 37–46.
11. *Kulikov V.V.* Kak my delali «AU!» // *Komp'yuterra.* 1998. №28 (257).
12. *Kulikov V.V., Polak Y.E.* Katalog russkoyazychnyh resursov interneta «AU!» // *Telematika'98.* SPb.: 1998. S. 334–335.

13. *Polak Y.*. Information search in Russian Internet on the eve of 2000 // Internet: Technologies and Services. The 1st IEEE International Conference and Exhibition, 1999.

14. *Ranganatan S.R.* Klassifikaciya dvoetochiem. Osnovnaya klassifikaciya. Per. s angl. pod. red. T.S. Gomolickoj i dr. M.: GPNTB SSSR, 1970.

15. *Foskett A.C.* The future of faceted classification // The future of classification, ed. by R. Marcella and A. Maltby, Aldershot: Gower, 2000. P. 69–80.

16. *Parakhina O.V.* Katalogi internet-resursov i ih mesto sredi poiskovyh instrumentov // Materialy konferencii predstavitelej regional'nyh nauchno-obrazovatel'nyh setej «RELARN-2001». SPb, 2001. S. 93–96.

17. *Kogalovskij M.R., Nevolin I.V., Parinov S.I.* Razvitie nauchnyh kommunikacij kak uslovie modernizacii ocenki nauchnoj rezul'tativnosti // Epistemologiya i filosofiya nauki. 2017. T. 51, № 1. S. 188–205

18. *Parinov S., Lyapunov V., Puzyrev R.* Sistema Socionet kak platforma dlya razrabotki nauchnyh informacionnyh resursov i onlajnovykh servisov // Elektronnye biblioteki. T. 6, № 1 (2003). URL: <https://elbib.ru/article/view/222>

19. *Parinov S.I.* Onlajnovoe budushchee nauki: naukometriceskaya signal'naya sistema: Preprint WP2/2007/01. M.: Izdatel'skij dom GU VSHE, 2007. 52 s.

20. *Polak Y.E., Sigalov A.V.* Internet v Rossii: setevaya infrastruktura i informacionnye resursy // Bibliotechnye komp'yuternye seti: Rossiya i Zapad. Sovremennye tendencii korporativnoj raboty bibliotek v setyah peredachi dannyh. Sbornik statej. Moskva, 1998. S. 10–34.

21. *Akinshin A.A., Polak Y.E.* Vklad CEMI v stanovlenie runeta. Informacionnye proekty // Vestnik CEMI RAN. 2022. T. 5, Vypusk 2. URL: <https://cemi.jes.su/s265838870019963-2-1>

СВЕДЕНИЯ ОБ АВТОРЕ



ПОЛЯК Юрий Евгеньевич – кандидат экономических наук, ведущий научный сотрудник Центрального экономико-математического института РАН (Москва). Подробнее: <http://computer-museum.ru/articles/sovets-muzeya/561/>

Yuri Evgenievich POLAK – Candidate of Economic Sciences, Leading Researcher, Central Economics and Mathematics Institute. Moscow, Russia. More detailed: <http://computer-museum.ru/articles/sovets-muzeya/561/>

email: polak@cemi.rssi.ru

ORCID 0000-0001-8411-335X

Материал поступил в редакцию 10 декабря 2022 года

ЦИФРОВАЯ ПЛАТФОРМА ДЛЯ СУПЕРКОМПЬЮТЕРНОГО МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРОЦЕССОВ НАПЫЛЕНИЯ

Н. И. Тарасов¹ [0000-0002-2342-1292], В. О. Подрыга² [0000-0001-7874-6978],

С. В. Поляков³ [0000-0003-1859-9034], А. В. Тимаков⁴ [0000-0001-6937-8018]

¹⁻⁴ *Институт прикладной математики им. М.В. Келдыша РАН, Миусская пл., 4, г. Москва, 125047*

¹nikita_tarasov@imamod.ru, ²pvictoria@list.ru, ³polyakov@imamod.ru,

⁴aletimakov@yandex.ru

Аннотация

Представлена цифровая платформа суперкомпьютерного моделирования в задачах напыления частиц на подложки. Целью работы является обсуждение общей архитектуры, технологического стека и особенностей реализации пользовательского интерфейса платформы. В ее основу положены веб-технологии доступа и управления расчетами, позволяющие реализовать пользовательскую систему для проведения полного цикла вычислительного эксперимента, включающего конфигурацию прикладных приложений, их запуск на удаленных вычислительных ресурсах, мониторинг прохождения задач, анализ и интерактивную визуализацию результатов. Взаимодействие пользователя с вычислительными ресурсами реализовано посредством графического интерфейса, не требующего от компьютера пользователя наличия дополнительного программного обеспечения, кроме веб-браузера. Важным преимуществом платформы является возможность проведения широкомасштабных компьютерных исследований в многопользовательском режиме, вытекающая из естественных принципов построения клиент-серверных приложений. Представляемая цифровая веб-платформа была успешно апробирована на вычислительных кластерах ИПМ им. М.В. Келдыша РАН при решении ряда актуальных математических задач нанотехнологии. Также с ее помощью последние 3 года проводится групповое обучение студентов МФТИ современным информационным технологиям.

Ключевые слова: суперкомпьютерное моделирование, цифровая платформа, веб-интерфейс, газодинамическое напыление частиц

ВВЕДЕНИЕ

Проблема напыления частиц на подложки входит в круг фундаментальных и прикладных задач нанотехнологии. В ее рамках актуально исследование математическими методами всех стадий процесса напыления, что позволит эффективно и экономно изготавливать наноматериалы и нанопокрyтия с заданными свойствами, используемые во многих отраслях промышленности. Вычислительное ядро, связанное с реализацией численных методов решения возникающих здесь математических задач, ориентировано изначально на применение мощных вычислительных кластеров и суперкомпьютеров. Оно предполагает работу исследователя с большими распределенными наборами данных, обработка которых также требует применения параллельных вычислений. В связи с этим целью настоящей работы является разработка специализированной веб-платформы для суперкомпьютерного моделирования процессов напыления в рамках различных физико-технических подходов.

Применение суперкомпьютерного моделирования в подобных исследованиях стало уже повсеместной практикой. Однако для достижения максимального эффекта от использования суперЭВМ необходима разработка многозадачных многопользовательских сред и сервисов. Последние в настоящее время разрабатываются в рамках облачных и веб-технологий. Практическая реализация таких сред предполагает создание удобного сетевого интерфейса пользователей и программных ядер, отвечающих за решение отдельных вычислительно емких задач прикладной направленности. Особое внимание при этом уделяется совместной работе пользователей над одним или несколькими проектами. В последнее время подобные программные среды получили название цифровых платформ, что связано с внедрением в них технологий искусственного интеллекта. В классическом CFD-моделировании эти технологии используются на этапе анализа полученных численных результатов большого объема, то есть больших данных.

АРХИТЕКТУРА ВЕБ-ПЛАТФОРМЫ

Выбирая концепцию создаваемой платформы, мы остановились на веб-решениях, которые в отличие от облачных технологий (имеется ввиду их вариант в виде инфраструктуры как услуги – IaaS) менее чувствительны к техническим параметрам конкретных компьютерных и суперкомпьютерных систем, менее затратные по срокам реализации и стоимости поддержки, но при этом поддерживают многоплатформенный и многопользовательский режимы работы. К тому же веб-технологии не исключают возможности использования систем облачных вычислений при наличии стандартного или специализированного интерфейса доступа и становятся некоторой единой точкой взаимодействия для конечного пользователя. При этом допускается использование кластеров и суперкомпьютеров, не интегрированных в облачные системы (таких, например, как гибридные суперкомпьютеры ИПМ им. М.В. Келдыша РАН К-60 или К-100 [1]).

Выполнение конкретных расчетных задач традиционно было предложено реализовать с помощью запуска приложений в пакетном режиме. Такой подход позволяет гибко настраивать работу конкретных расчетных программ посредством формирования необходимых конфигурационных файлов и скриптов, которые можно запускать в локальном или удаленном режимах с помощью ssh-консоли. Подобная технология, реализуемая непосредственно пользователем, требует от него определенных знаний и навыков работы с терминалом, а также самостоятельного учета, мониторинга и анализа всех запущенных или завершившихся расчетов. Последнее фактически делает работу пользователя неэффективной, а иногда и просто невозможной. Для решения этой технической проблемы и создается интерактивная и интеллектуальная в определенном смысле веб-среда (веб-платформа), которая принимает на себя максимальное количество административных, технических и прочих трудоемких функций, а также предоставляет пользователю дружественное визуальное общение с большим количеством программ, данных и коллег по проектам.

Предлагаемая нами цифровая веб-платформа, схема которой изображена на рис. 1, представляет собой универсальный инструмент взаимодействия разработчиков и пользователей прикладных программных кодов с доступными им аппаратными средствами (от персональных компьютеров до рабочих станций и

мощных вычислительных кластеров) посредством графического веб-интерфейса, поддерживающего следующие возможности:

- авторизация пользователей;
- проведение полного цикла компьютерного моделирования процесса напыления частиц, в том числе многопользовательского;
- возможность запуска заданий на удаленных суперкомпьютерных ресурсах;
- формирование базы данных расчетов и обеспечение доступа к их результатам;
- обеспечение возможностей анализа результатов.

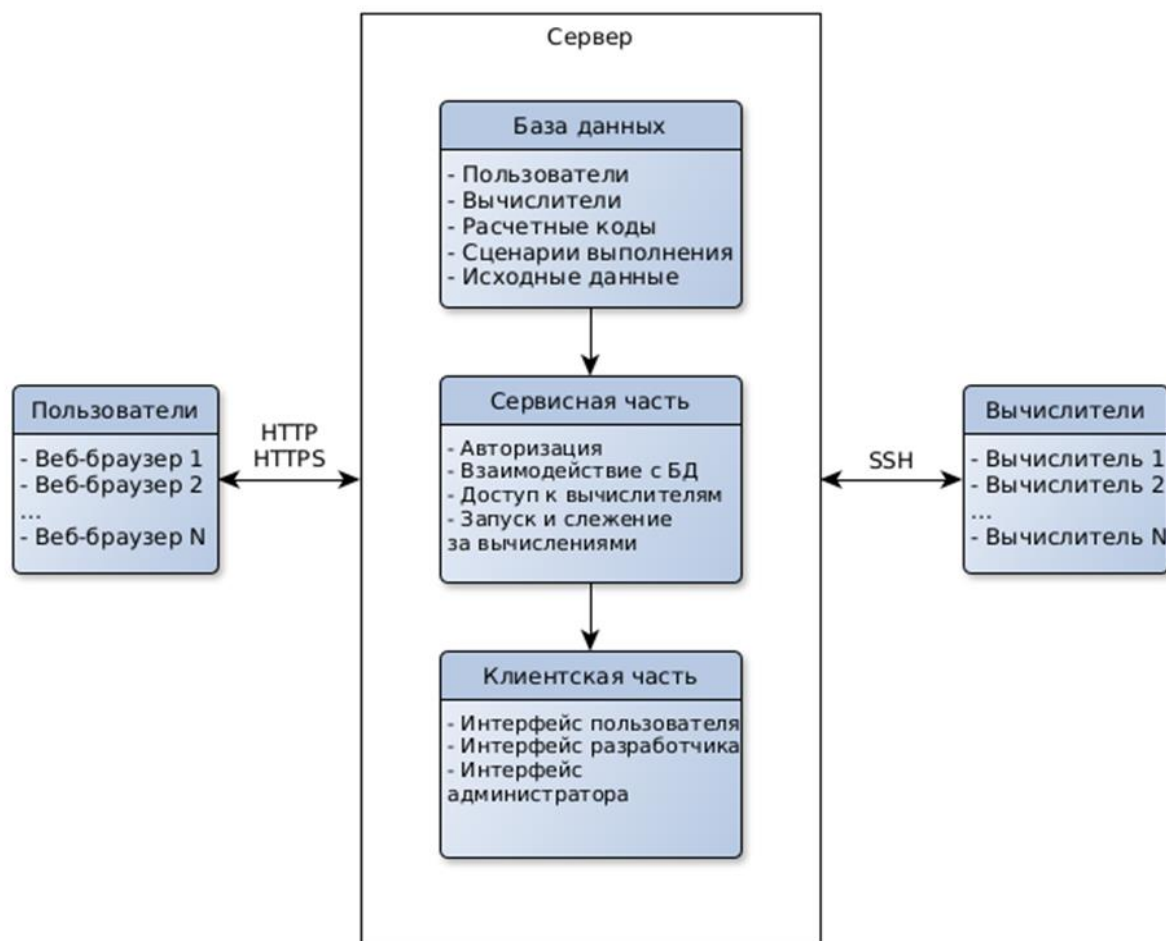


Рис. 1. Схема цифровой веб-платформы.

Взаимодействие пользователей между собой заключается в следующем. Во-первых, это может быть совместная разработка кодов приложений, для чего используются система контроля версий и общие папки на сервере проекта. Во-

вторых, это совместное проведение расчетов и обработка их результатов в рамках общего проекта. Такая возможность регулируется с помощью гибкой системы прав доступа к проекту и защиты от одновременного запуска расчетных программ с одинаковыми начальными данными из разных логинов. Отдельно отметим особую актуальность данного функционала, вытекающую из практики проведения комплексных вычислительных экспериментов в выбранной предметной области. Применяемый здесь многомасштабный подход при моделировании сложных физических процессов в широком диапазоне пространственных размеров (от метров до нанометров) опирается на синтез классических моделей механики сплошных сред, методов частиц и/или Монте-Карло и численный анализ свойств веществ и материалов на их основе. Очевидно, проведение вычислительного эксперимента подобного объема и качества силами одного специалиста представляет собой трудоемкую задачу. Однако при наличии возможности участия нескольких экспертов из различных прикладных областей, а также при использовании накопленных баз данных по свойствам веществ и материалов и решениям отдельных актуальных подзадач проведение комплексных исследований в рамках выполнения совместных расчетных проектов становится достижимым и весьма эффективным решением.

НАБОР ТЕХНОЛОГИЙ

Первая разработка предлагаемой цифровой платформы [2] основывалась на Django [3] – библиотеке для языка программирования Python, содержащей все необходимое для создания клиент-серверного решения, в том числе механизм объектно-реляционного отображения (ORM) для взаимодействия с базами данных, генератор веб-страниц из html-шаблонов, системы маршрутизации и авторизации. В процессе разработки выяснилось, что использования подхода шаблонных страниц недостаточно для создания интерактивного пользовательского интерфейса, обеспечивающего возможности, необходимые для проведения вычислительного эксперимента. Для поддержки требуемого функционала приходилось использовать JavaScript код, исполняемый непосредственно в веб-браузере. В итоге от использования шаблонов было решено отказаться в пользу разработки

одностраничного приложения (SPA) в реактивном стиле, для чего была применена библиотека Vue.js [4]. В данной разработке было решено исключить также и Django, поскольку большая часть его функционала, заключающаяся в шаблонах веб-страниц, перестала использоваться.

Программной основой для серверной части цифровой платформы стал Node.js [5], являющийся расширением JavaScript. Обычно язык программирования JavaScript используется на стороне веб-клиента и поддерживается во всех современных браузерах. Однако он имеет существенные ограничения по функциональности, в том числе невозможность взаимодействия с устройствами ввода-вывода и системными потоками напрямую. Для расширения JavaScript до языка общего назначения в целях создания серверного решения и используется Node.js.

Применение Node.js позволило, с одной стороны, добиться однородности кодовой базы между клиентской и серверной частями, а с другой – использовать высокую степень асинхронности JavaScript на сервере. Это особенно полезно для целей слежения за прохождением задач на удаленных вычислителях, выполнения длительных заданий и масштабирования системы.

Для обеспечения маршрутизации, а также использования плагинов безопасности был использован фреймворк Express.js [6], авторизация реализована посредством библиотеки passport.js [7]. Взаимодействие с базой данных осуществлялось с помощью библиотеки ORM Sequelize [8]. В качестве базы данных на стадии разработки использовалась SQLite [9].

Поскольку создаваемая цифровая платформа является закрытой для общего доступа, было решено организовать клиент-серверное взаимодействие на основе веб-сокетов [10]. Веб-сокеты представляют собой надстройку над протоколом HTTP. В отличие от исходного протокола они поддерживают соединение после начального клиентского запроса о подключении. При успешном установлении такового клиент и сервер могут произвольно обмениваться сообщениями. Сообщения, отправляемые посредством веб-сокета, включают в себя исключительно тело запроса, а HTTP-заголовок повторно не отправляется, что позволяет существенно оптимизировать клиент-серверное общение. Для организации веб-сокетов на сервере применялась библиотека ws [11]. Обеспечение HTTPS-протокола достигалось посредством обратного прокси сервера nginx[12].

Для использования конкретных расчетных приложений и их объединения в функциональные группы (сценарии) в качестве формата метаописания используются файлы текстового формата YAML [13]. Выбранный формат является, с одной стороны, удобным для пользовательского ручного формирования, с другой – поддерживается в стандартной библиотеке JavaScript для сериализации (преобразование разнородных данных в битовую последовательность) и десериализации (обратное преобразование).

Клиентская часть цифровой платформы была реализована с применением реактивной парадигмы, организованной с помощью использования фреймворка Vue.js и библиотеки компонентов Quasar [14]. Для SPA-маршрутизации применялось естественное в данном случае решение Vue-router [15], для организации централизованного хранилища данных использовалась Vuex [16] (хранение информации о пользователе, открытого веб-сокета, уведомлений). В отношении компонента анализа результатов расчетов использован ParaViewWeb [17, 18] с встраиванием клиентской части в реализации Visualizer.js [19]. Данное решение обеспечивает поддержку форматов VTK, возможность интерактивной серверной или клиентской отрисовки, в зависимости от аппаратного обеспечения на компьютере пользователя. ParaViewWeb поддерживает большое количество фильтров данных, сопоставимое со стандартным «толстым» приложением ParaView. Отдельным преимуществом ParaViewWeb является возможность реализации собственного клиента на стандартном стеке цифровой платформы, что может в будущем привести к отказу от встраивания клиентской части платформы в Visualizer.js.

СЕРВЕР ЦИФРОВОЙ ПЛАТФОРМЫ

Взаимодействие с сервером цифровой платформы основано на использовании графических и иных элементов пользовательского интерфейса, в результате взаимодействия с которыми веб-браузер отправляет на сервер соответствующий запрос. Маршрутизация запросов осуществляется посредством унифицированных указателей ресурса (URL) и имеет следующий вид:

- GET /auth/ – проверка актуальности сессии текущего пользователя;
- POST /auth/login/ – авторизация пользователя;
- POST /auth/logout/ – завершение сессии текущего пользователя;

- POST /auth/register/ – регистрация нового пользователя;
- GET /files/ – получение файла, расположенного в рабочей директории веб-сервера;
- GET /files/ssh/ – получение файла с удаленного вычислительного ресурса;
- POST /files/tmp/ – отправка пользовательского файла на сервер во временное хранилище;
- GET /api/ – при обращении открывает основное веб-сокет соединение, предназначенное для обеспечения взаимодействия между клиентом (веб-браузером) и сервером цифровой платформы;
- GET /terminal/ - при обращении открывает веб-сокет соединение с веб-браузером пользователя, а также устанавливает ssh-соединение с требуемым удаленным вычислителем. После успешной установки соединений сервер связывает веб-сокет с ssh-клиентом.

Как указано выше, основным каналом взаимодействия между веб-браузером компьютера пользователя и сервером цифровой платформы является веб-сокет, открываемый при обращении по URL-адресу /api/. Сообщения, отправляемые клиентом, должны содержать обязательное поле «url» для вызова корректного обработчика запроса и опциональное поле «data», содержащее добавочную информацию. Поле «url» должно представлять собой строку, во многом соответствующую стандартному URL-адресу. Ответ сервера отправляется в сообщении, содержащем поля «url», «body» и числовой код в поле «status».

Необходимость наличия в ответе поля «url» объясняется особенностью работы веб-сокета, заключающейся в отсутствии классической схемы: запрос клиента и последующий ответ сервера. При использовании соединения данного типа клиент может только лишь подписаться на сообщения, отправляемые сервером (установка обработчика onmessage). В рамках разработанного подхода единый обработчик входящих сообщений вызывает конкретные клиентские функции, установленные при отправке запроса на сервер, на основании поля «url» ответа. Существуют два типа устанавливаемых обработчиков – постоянные (используемые при реагировании на сообщения о новом уведомлении или сообщении, подписке на изменение статуса расчетного задания) и единоразовые (используемые

при запросе информации из базы данных, списка файлов на удаленном вычислителе).

Отметим, что масштабирование веб-сервера производится посредством создания подпроцессов средствами Node.js. При этом главный процесс обеспечивает задачи слежения за расчетными заданиями и состоянием вычислительных ресурсов. Процессы-потомки используются для обработки пользовательских запросов и обслуживания установленных веб-сокет соединений.

Для хранения и обеспечения пользовательской информации используется база данных, имеющая структуру, изображенную на рис. 2.

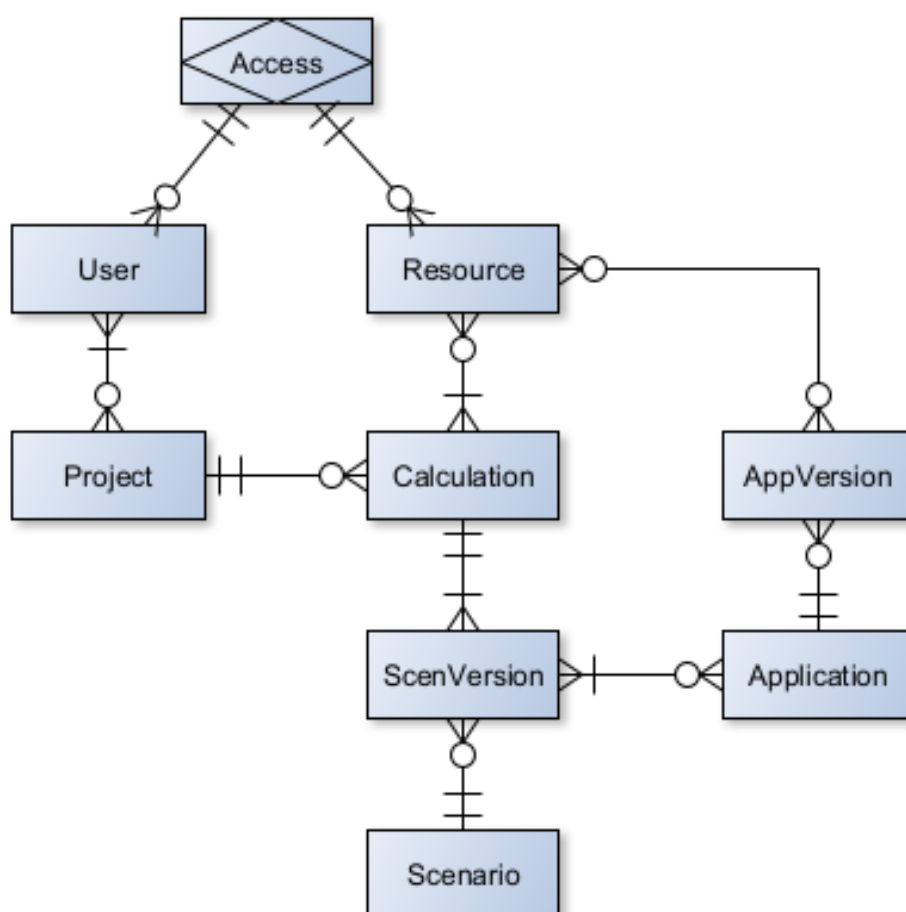


Рис. 2. Схема связей между компонентами базы данных.

Для поддержки возможности динамической регистрации удаленного вычислительного ресурса необходима информация об адресе вычислительного ресурса, порте ssh-сервера, системе прохождения задач, возможностях задания переменных окружения. Эта информация размещается в базе данных цифровой

платформы в таблице «Resource». Для осуществления ssh-доступа к удаленному вычислителю и разделения на нем пользователей используется таблица «Access», содержащая ссылки на записи пользователя системы и вычислителя, а также дополнительные переменные окружения и указание на используемые пользователем версии приложений. Подключение к удаленному вычислителю производится с использованием RSA-ключа веб-системы.

Описание расчетных приложений, помимо информационного, включает в себя также задание шаблонного метафайла, содержащего инструкции сборки, запуска, описание входных и выходных параметров, хранение которого производится в таблице «Application». Поддержка многообразия версий приложения достигается с помощью таблицы «AppVersion», содержащей дополнительно информацию о пути к исходному коду конкретной версии и итоговом метафайле, используемом для генерации графического пользовательского интерфейса.

Сценарии (таблица «Scenario») используются для объединения нескольких приложений в единый расчетный цикл. Они обеспечивают возможность запуска приложений и связывания их входных и выходных данных (с помощью файлового интерфейса и/или командной строки). Для описания сценариев используется файл определенной структуры. Поддержка многообразия версий сценариев производится при помощи таблицы «ScenVersion». Итоговое метаописание сценария используется при генерации графического пользовательского интерфейса для настройки прикладного ПО.

Суть каждого проекта платформы носит информационный характер, располагается в таблице «Project» и используется для объединения в единую смысловую конструкцию нескольких расчетов. Расчеты представляют собой циклы и результаты вычислений. Информация о них хранится в таблице «Calculation». Расчеты с точки зрения системы управления платформой являются основным способом взаимодействия с конкретным удаленным вычислителем. При этом под вычислителем понимается не только удаленный компьютер или кластер, но и конфигурация его использования (например, количество удаленных процессов, участвующих в расчете). При настройке конкретного расчета платформа генери-

рует графический интерфейс пользователя для ввода исходных данных, управления расчетным заданием (запуск, останов, продолжение, реконфигурация) и доступа к выходным данным на удаленном вычислителе.

ПОЛЬЗОВАТЕЛЬСКИЙ ВЕБ-ИНТЕРФЕЙС

Иллюстрацией реализации совместных расчетных проектов могут служить исследования ИПМ им. М.В. Келдыша РАН, ОИЯИ и ИПЛИТ РАН в рамках разработки программного комплекса GIMM_NANO [20]. Однако в этом решении еще не были задействованы в современном варианте ни облачные, ни веб-технологии. Тем не менее, объединение научных разработок вышеуказанных партнеров привело к существенному повышению качества научных исследований в области перспективных нанотехнологий.

Перейдем теперь к описанию работы пользователей в рамках предлагаемой цифровой платформы. После успешной авторизации пользователь, в соответствии с его ролью (администратор, разработчик или обычный пользователь), получает доступ к разделам системы, которые можно разбить на следующие функциональные группы:

- вычислители, доступные администраторам (см. рис. 3);
- приложения и сценарии, формируемые разработчиками (см. рис. 4);
- проекты и расчеты, используемые обычными пользователями (см. рис. 5).

Подобное распределение пользователей по ролям видится необходимым как для разграничения областей ответственности, так и для обеспечения максимально эффективной их работы в соответствии со специализацией. Так, основной деятельностью обычного пользователя являются проведение вычислений и анализ результатов компьютерного моделирования. При этом непосредственный доступ пользователя к кодам приложений, сценариям их работы и конкретным вычислителям ограничен возможностью получения справочной информации. Разработчики программного обеспечения заняты созданием новых и модификацией имеющихся приложений и сценариев, а также используют доступные вычислительные ресурсы с целью тестирования и адаптации кодов к тем или иным компьютерным архитектурам. Администраторы системы управляют аккаунтами поль-

зователей, регистрируют в системе новые вычислительные ресурсы, подготавливают их к проведению пользователями масштабных вычислительных экспериментов, обеспечивают целостность и безопасность платформы в целом. Если все три роли будут совмещаться «в одних руках», неизбежно начнутся проблемы деградации платформы ввиду нехватки нужных компетенций и времени на оперативное реагирование на возникающие нештатные ситуации.

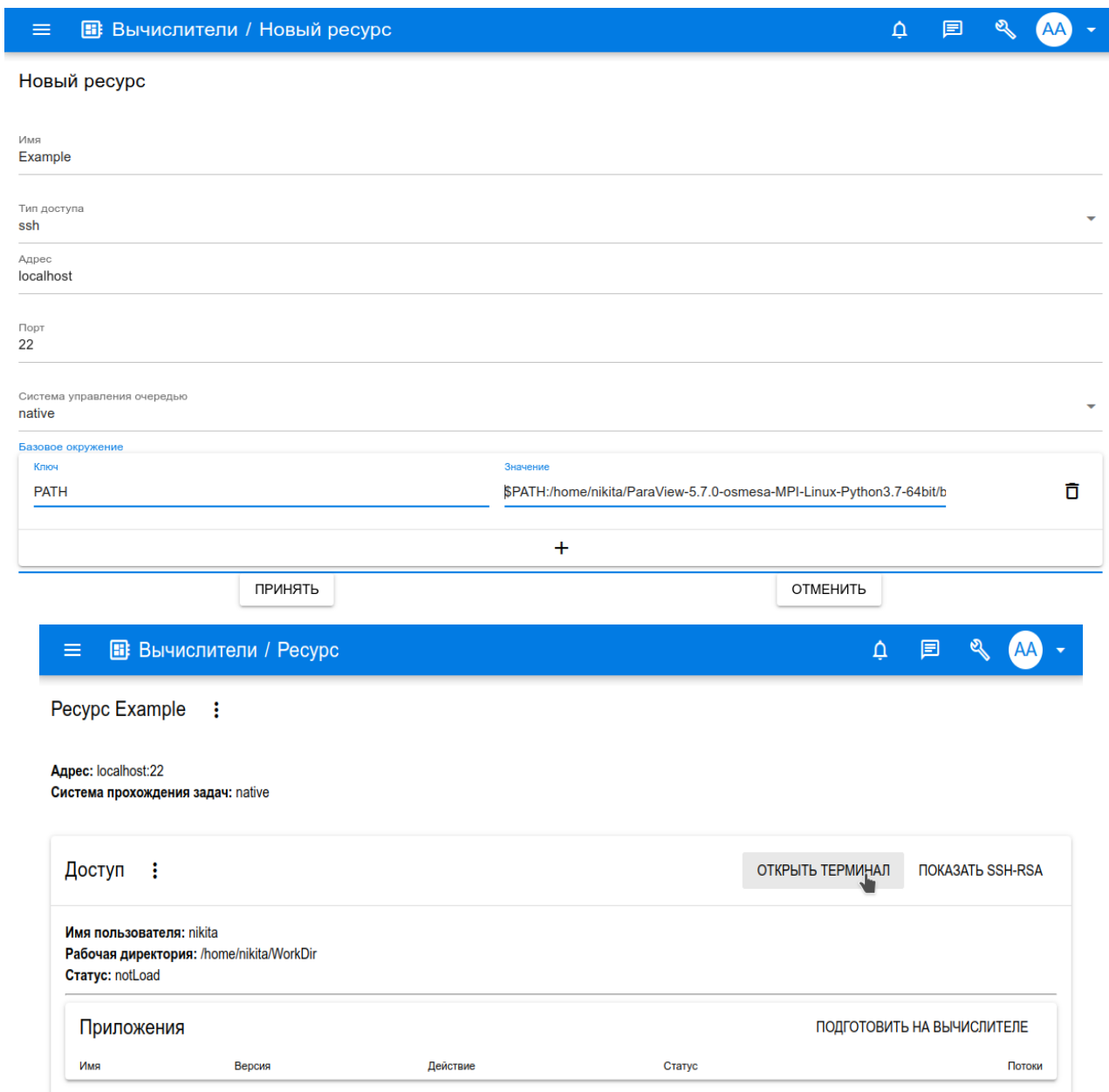


Рис. 3. Интерфейс администратора взаимодействия с вычислителями.

The screenshot displays the interface for developing and preparing applications. It is divided into two main sections.

Top Section: Passport Template Editor

The top part shows a code editor for a 'Паспорт' (Passport) template. The code is as follows:

```

1 build:
2   mkdir build; cd build; cmake ../; make -j4
3 exec:
4   comm: build/solver config.json
5 runBy:
6   type: exec
7 inputs:
8   - type: file
9     ident: mesh
10    name: Mesh file
11    descr: Mesh file in inner format
12   options:
13     format: data
14     path: mesh.txt
15   - type: file
16     ident: jsoncfg
17     name: Configuration
18     descr: Configuration file in json format
19   options:
20     format: tabs
21     vertical: true
22     fType: json
23     path: config1.json
24   childs:
25   - type: category
26     ident: main
27     name: Main
28     descr: Main global parameters
29   options:
30     format: form
31   childs:

```

Below the code editor is a progress bar for the source code upload, showing 'Исходный код' (Source code) at 111.9KB / 100.00%. A file named 'src.tar.gz' is listed as uploaded.

Bottom Section: Application Management

The bottom part shows the 'Приложение Solver' (Application Solver) management interface. It includes the following information:

- Идентификатор: solver
- Автор: Developer Developer
- Публичный: Да

There are three tabs: ОПИСАНИЕ, ШАБЛОН ПАСПОРТА, and ПОДГОТОВКА. The 'ПОДГОТОВКА' (Preparation) tab is active, showing a 'Ресурсы' (Resources) table with a 'ПОДГОТОВИТЬ НА ВЫЧИСЛИТЕЛЯХ' (Prepare on calculators) button.

Имя	Версия	Действие	Статус	Потоки
Example	V1	build	success	STDOUT STDERR

On the right, the 'Версии' (Versions) table is shown with a search bar and a 'НОВАЯ ЗАПИСЬ' (New Record) button.

Имя	Создан	Изменен	Действия
V1	04.09.2022, 13:26:04	04.09.2022, 13:26:04	🔗 ✎ 🗑

At the bottom right, it indicates 'Строк на странице: 5' (Rows per page: 5) and '1-1 из 1' (1 of 1).

Рис. 4. Интерфейс разработчика регистрации и подготовки приложений.

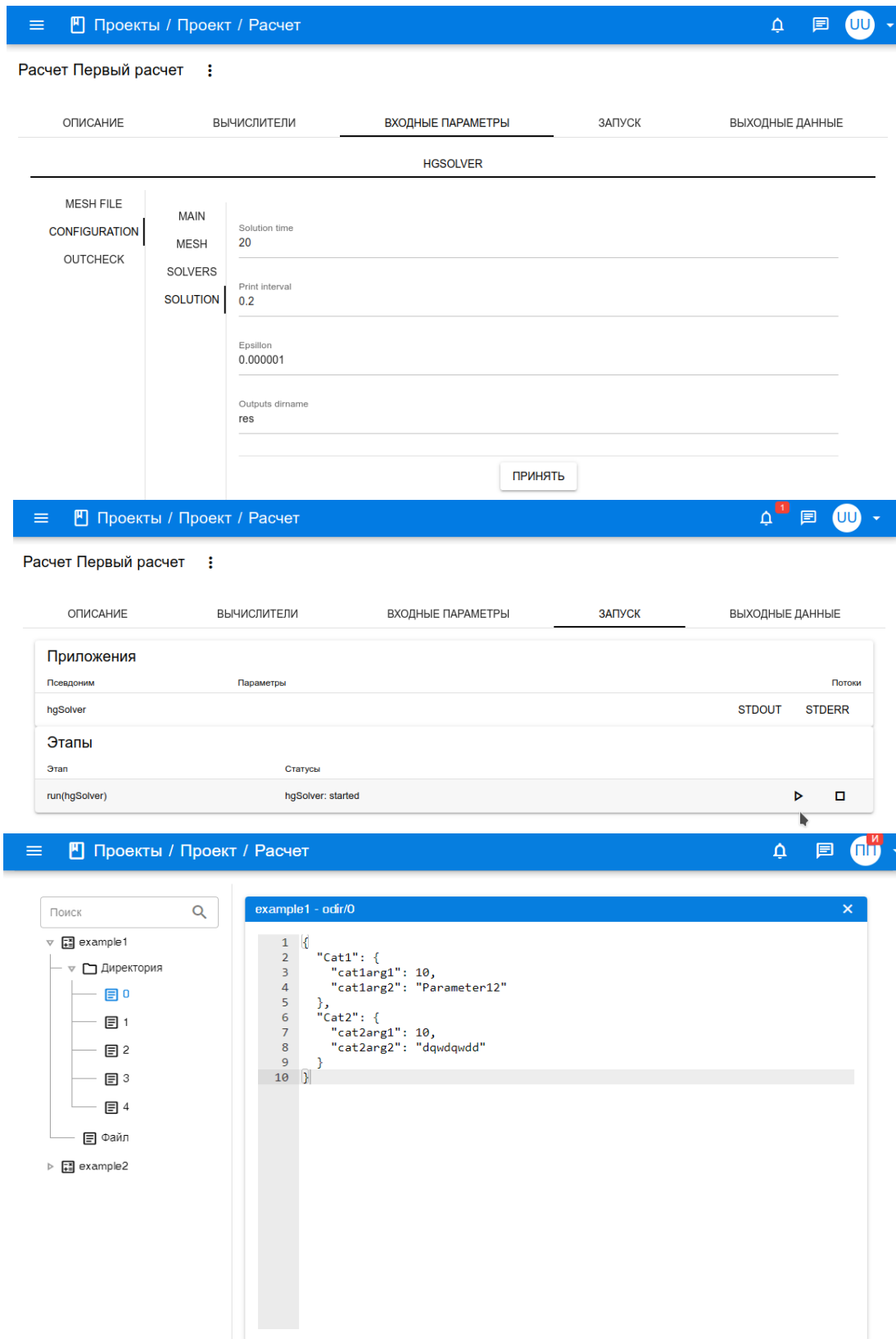


Рис. 5. Пользовательский интерфейс взаимодействия с расчетами.

Основой численного эксперимента в рамках платформы является расчет, при доступе к странице которого пользователю предоставляются графические формы выбора вычислительных ресурсов и задания исходных данных, компонента запуска и слежения, список результатов моделирования. Генерация интерфейса происходит в соответствии с выбранным сценарием на основе метаописания – текстового файла, содержащего список используемых приложений, а также связывание управляющих параметров, файлов данных, логику и порядок исполнения. Расчетные приложения платформы имеют метаописания, включающие в себя команды сборки и запуска, списки управляющих параметров, исходных данных и результатов исполнения.

Создание новых расчетов осуществляется на странице конкретного проекта, который позволяет каталогизировать и объединить некоторое их количество, задать общее описание.

В качестве примера работы первой версии веб-платформы приведем исследования в рамках проекта «Напыление». В его рамках возможно проведение расчетов нескольких прикладных задач. Одной из них является сверхзвуковое газодинамическое напыление отдельных металлических нанокластеров на подложку. В наших предыдущих работах по этой теме (см., например, [21, 22]) были представлены модель, численные алгоритмы, некоторые результаты конкретных параллельных вычислений и их анализа. Однако для выполнения всех вышеуказанных действий необходимо было программно реализовать сложную цепочку вычислительных экспериментов. Она начиналась с генерации исходных микро- и нанообъектов – например, столба азота микронных размеров и никелевых нанокластеров, находящихся в состоянии термодинамического равновесия. На основе этих данных генерировалось стартовое состояние системы газ–металл, в которой нанокластеры погружены в неподвижный азот. Затем производились разгон газовой среды и формирование ударной волны, которая заставляла кластеры двигаться в сторону подложки с нужной скоростью. На всех этапах такого моделирования использовались длительные параллельные вычисления. Переход от одного вычислительного эксперимента к другому производился пользователем «вручную». В частности, все расчеты сохранялись на дисках разных суперкомпьютеров, и их продолжение на очередной стадии требовало от пользователя переноса

больших данных с одной вычислительной системы на другую. Большой проблемой было отслеживание качества решения на всех стадиях вычислений ввиду отсутствия универсальных средств постобработки и визуализации распределенных данных большого объема. Частично эти и другие задачи были решены с помощью программных средств, разработанных нами ранее (см., [23, 24]). Однако полный цикл вычислений с минимальным участием пользователя стал доступным только в рамках разработки представляемой цифровой веб-платформы.

В частности, хотелось бы отметить в качестве нового программного решения визуализацию распределенных данных сверхбольшого объема. Оно стало возможным с появлением пакета ParaViewWeb, который пока не стал традиционным средством визуального анализа в отличие от его классической параллельной версии. В рамках реализации представляемой веб-платформы удалось интегрировать этот пакет в общую систему управления платформы и, тем самым, обеспечить максимальную полноту его функционала и управляемость со стороны пользователя.

В качестве примера на рис. 6 представлен этап визуализации рассчитанных на вычислительном кластере данных, представляющих собой распределения плотности газовой фазы в начальный и конечный моменты времени, описывающие процесс напыления на подложку.

Наиболее значимым в данном примере является то, что такая возможность визуализации позволяет анализировать результаты во время расчета на выбранном суперкомпьютере (например, можно контролировать состояние кластера и его маршрут, проверять целостность, изменение температуры, формы и т. д.). Для этого требуется только доступ системы управления платформой к дисковому пространству суперкомпьютера, на котором сохраняются результаты вычислений.

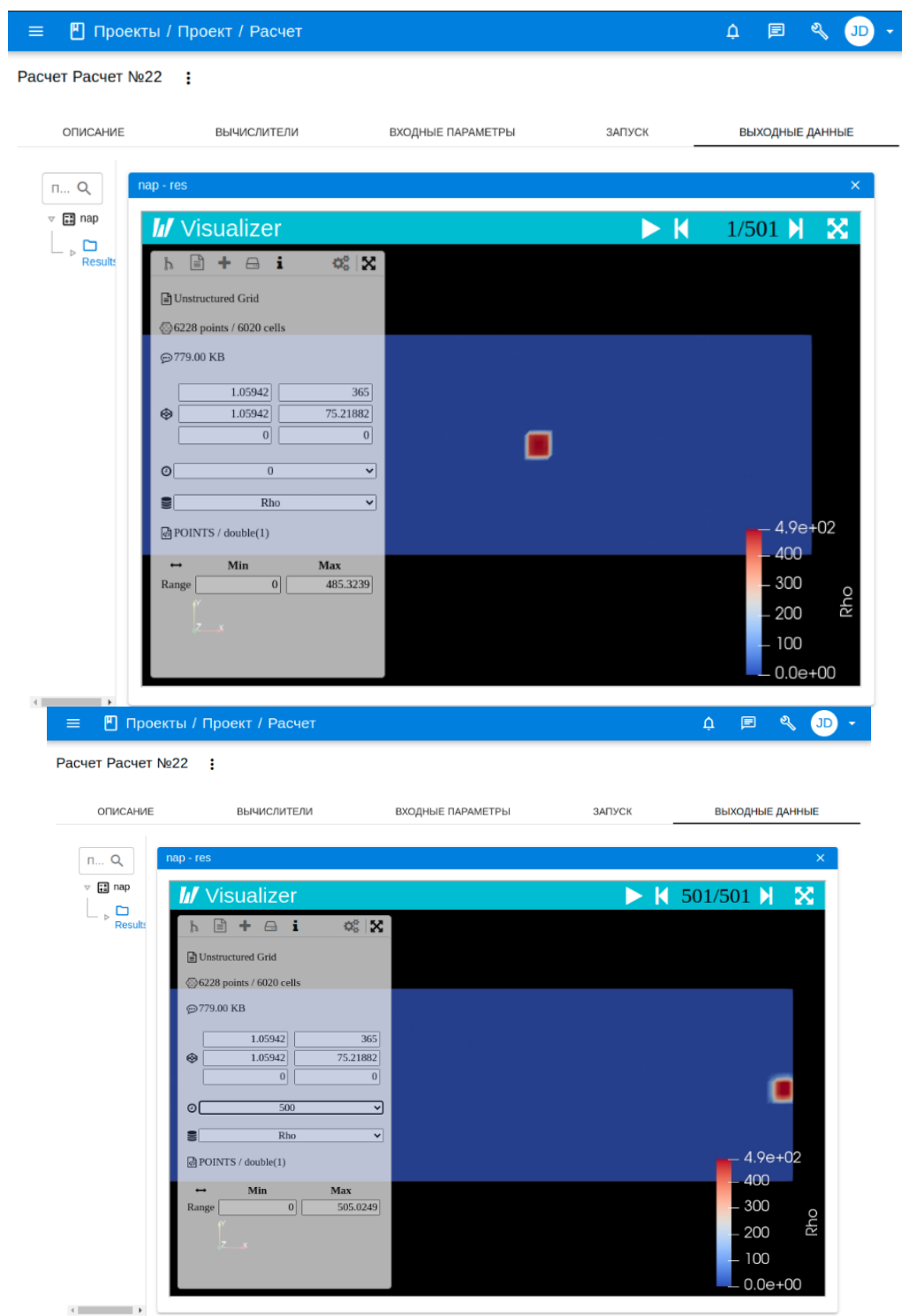


Рис. 6. Пример визуализации в рамках цифровой платформы.

Реализованное программное решение было апробировано на кластерах ИПМ им. М.В. Келдыша РАН при решении ряда задач вычислительной газо- и гидродинамики [25–28]. Эффективность его работы была подтверждена в период

пандемии COVID-19, когда практически все участники проекта (порядка 10 человек) работали удаленно. Также с помощью разработанной учебной версии веб-платформы уже третий год реализуется проект, связанный с групповым обучением студентов МФТИ современным информационным технологиям (количество студентов в группах 12–15 вместе с преподавателем). При этом ограничением производительности веб-платформы является конкретный ресурс задействованных серверов проекта. Например, обычный веб-сервер легко справляется с одновременной работой 100 и более пользователей, выполняя одновременно до 30000 запросов. В представляемом решении основная нагрузка приходится не столько на обработку веб-запросов, сколько на сценарии работы прикладных сервисов, связанных с пред- и постобработкой больших данных. Поскольку все эти задачи выполняются на удаленных вычислительных ресурсах, то большая часть нагрузки с клиентской части компьютера пользователя и веб-сервера управления платформой снимается. Исключением пока является просмотр множества видеопотоков, инициированных несколькими пользователями.

ЗАКЛЮЧЕНИЕ

Подводя итог, отметим, что на основе доработки ранее созданных пробных решений удалось реализовать новую версию веб-платформы, предназначенную для компьютерного и суперкомпьютерного моделирования задач напыления наночастиц на подложки и других связанных задач нанотехнологии. С ее помощью проведен ряд вычислительных экспериментов по напылению никелевых кластеров на подложки. В качестве суперкомпьютерных систем при этом были использованы гибридные суперкомпьютеры К60 и К100 и специализированные серверы ЦКП ИПМ им. М.В. Келдыша РАН. Количество пользователей цифровой платформы с учетом студенческой аудитории составило 25 чел. Отказов цифровой платформы при одновременной работе всех вышеуказанных пользователей не зафиксировано. В перспективе планируются расширение базы приложений, поддерживаемых платформой, и расширение функционала со стороны интерфейса пользователя. Планируется также поддержка альтернативных систем управления расчетными заданиями на удаленных вычислительных системах.

Благодарности

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 21-71-20054).

СПИСОК ЛИТЕРАТУРЫ

1. Center of Collective Usage of KIAM RAS. URL: <https://ckp.kiam.ru/?home>
2. *Puzyrkov D.V., Podryga V.O., Polyakov S.V.* Cloud service for HPC management: ideas and appliance // *Lobachevskii Journal of Mathematics*. 2018. V. 39, No. 9. P. 1251–1261. <https://doi.org/10.1134/S1995080218090172>
3. The web framework for perfectionists with deadlines | Django. URL: <https://www.djangoproject.com/>
4. Vue.js–The Progressive JavaScript Framework | Vue.js. URL: <https://vuejs.org/>
5. About | Node.js. URL: <https://nodejs.org/en/about/>
6. Express – Node.js web application framework. URL: <https://expressjs.com/>
7. Passport.js. URL: <https://www.passportjs.org/>
8. Sequelize | Feature-rich ORM for modern TypeScript & JavaScript. URL: <https://sequelize.org/>
9. SQLite Home Page. URL: <https://sqlite.org/index.html>
10. RFC 6455 – The WebSocket Protocol. URL: <https://datatracker.ietf.org/doc/html/rfc6455>
11. ws: a Node.js WebSocket library. URL: <https://github.com/websockets/ws>
12. nginx. URL: <https://nginx.org/>
13. The Official YAML Web Site. URL: <https://yaml.org/>
14. Quasar Framework. URL: <https://quasar.dev/>
15. Home | Vue Router. URL: <https://router.vuejs.org/>
16. What is Vuex? | Vuex. URL: <https://vuex.vuejs.org/>
17. Web | ParaView. URL: <https://www.paraview.org/web/>
18. ParaViewWeb. URL: <http://kitware.github.io/paraviewweb/>
19. Visualizer. URL: <https://kitware.github.io/visualizer/>

20. Бондаренко А.А., Кононов Э.М., Косолапов О.А., Поляков С.В., Якобовский М.В. Программный комплекс GIMM_NANO // Научный сервис в сети Интернет: все грани параллелизма: Труды Международной суперкомпьютерной конференции (23–28 сентября 2013 г., г. Новороссийск). М.: Изд-во МГУ, 2013. С. 333–337. URL: <http://agora.guru.ru/abrau2013/pdf/333.pdf>

21. Поляков С.В., Подрыга В.О., Пузырьков Д.В., Кудряшова Т.А. Суперкомпьютерное молекулярное моделирование газодинамического напыления наночастиц на подложку // Суперкомпьютерные дни в России: Труды международной конференции (24–25 сентября 2018 г., г. Москва). М.: Изд-во МГУ, 2018. С. 782–792. URL: <http://russianscdays.org/files/pdf18/782.pdf>

22. Podryga V.O., Polyakov S.V. Multiscale mathematical modeling of the metal nanoparticles motion in a gas flow // In: Dimov I., Farago I., Vulkov L. (eds). Seventh International Conference on Finite Difference Methods. Theory and Applications. FDM 2018. Lecture Notes in Computer Science. Cham: Springer, 2019. V. 11386. P. 387–394. https://doi.org/10.1007/978-3-030-11539-5_44

23. Puzyrkov D., Podryga V., Polyakov S., Jakobovskii M. KIAM_JOB_CONTROL task management environment and its application to cloud and GRID computing) // Selected Papers of the 7th International Conference Distributed Computing and Grid-technologies in Science and Education (GRID 2016), ed. by V. Korenkov, T. Zaikina, F. Nechaevskiy, July 4–9, 2016, Dubna, Russia. CEUR Workshop Proceedings. 2017. V. 1787. P. 416–422. URL: <http://ceur-ws.org/Vol-1787/416-422-paper-72.pdf>

24. Kornilina M.A., Podryga V.O., Polyakov S.V., Puzyrkov D.V., Jakobovskiy M.V. Cloud service for solution of promising problems of nanotechnology // Supercomputing Frontiers and Innovations. 2017. V. 4, No. 4. P. 66–79. <https://doi.org/10.14529/jsfi170405>

25. Polyakov S.V., Podryga V.O., Kudryashova T.A. HPC simulation of non-linear processes in microsystems gas-metal // Lobachevskii Journal of Mathematics. 2020. V. 41, No. 8. P. 1554–1562. <https://doi.org/10.1134/S1995080220080168>

26. Podryga V.O., Polyakov S.V., Tarasov N.I. Developing of multiscale approach to HPC-simulation of multiphase fluid flows // Lobachevskii Journal of Mathematics. 2021. V. 42, No. 11. P. 2626–2636. <https://doi.org/10.1134/S1995080221110160>

27. Tarasov N., Karamzin Yu., Kudryashova T., Podryga V., Polyakov S. Computer modeling of air flows purification using sorption filters // Journal of Physics: Conf. Series. 2021. V. 2028. Paper 012025 (7 p.).

<https://doi.org/10.1088/1742-6596/2028/1/012025>

28. Kudryashova T.A., Polyakov S.V., Tarasov N.I. Mathematical modelling of electrophysical water treatment // Defect and Diffusion Forum. 2021. V. 412. P. 149–162.

<https://doi.org/10.4028/www.scientific.net/DDF.412.149>

DIGITAL PLATFORM FOR SUPERCOMPUTER MATHEMATICAL MODELING OF SPRAYING PROCESSES

N. I. Tarasov¹[0000-0002-2342-1292], V. O. Podryga²[0000-0001-7874-6978],

S. V. Polyakov³[0000-0003-1859-9034], A. V. Timakov⁴[0000-0001-6937-8018]

¹⁻⁴*Keldysh Institute of Applied Mathematics of RAS, Miusskaya sq., 4, Moscow, 125047*

¹nikita_tarasov@imamod.ru, ²pvictoria@list.ru, ³polyakov@imamod.ru,

⁴aletimakov@yandex.ru

Abstract

The work presents a digital platform for supercomputer modeling the problems of spraying the particles on substrates. The purpose of this work is to discuss the general architecture, technology stack and implementation features of the platform's user interface. The platform is based on web technologies for access and management of calculations, which allow implementing a user system for conducting a full cycle of a computational experiment, including the configuration of applied applications, their launch on remote computing resources, monitoring the completion of tasks, analysis and interactive visualization of results. User interaction with computing resources is implemented through the graphical interface that does not require the client computer to have any additional software, except actual version of a modern web browser. An important advantage of the platform is the ability to make large-scale computer research in a multi-user mode that is based on the natural principles of building client-server applications. The presented digital web platform was successfully tested on

computing clusters of the KIAM RAS in solving a number of the topical mathematical problems of nanotechnology. Also, with its help, for the last 3 years, group training of MIPT students in modern information technologies has been carried out.

Keywords: *supercomputer modeling, digital platform, web interface, gas-dynamic spraying of particles*

REFERENCES

1. Center of Collective Usage of KIAM RAS. URL: <https://ckp.kiam.ru/?home>
2. Puzyrkov D.V., Podryga V.O., Polyakov S.V. Cloud service for HPC management: ideas and appliance // Lobachevskii Journal of Mathematics. 2018. V. 39, No. 9. P. 1251–1261. <https://doi.org/10.1134/S1995080218090172>
3. The web framework for perfectionists with deadlines | Django.
URL: <https://www.djangoproject.com/>
4. Vue.js–The Progressive JavaScript Framework | Vue.js.
URL: <https://vuejs.org/>
5. About | Node.js. URL: <https://nodejs.org/en/about/>
6. Express – Node.js web application framework. URL: <https://expressjs.com/>
7. Passport.js. URL: <https://www.passportjs.org/>
8. Sequelize | Feature-rich ORM for modern TypeScript & JavaScript.
URL: <https://sequelize.org/>
9. SQLite Home Page. URL: <https://sqlite.org/index.html>
10. RFC 6455 – The WebSocket Protocol.
URL: <https://datatracker.ietf.org/doc/html/rfc6455>
11. ws: a Node.js WebSocket library. URL: <https://github.com/websockets/ws>
12. nginx. URL: <https://nginx.org/>
13. The Official YAML Web Site. URL: <https://yaml.org/>
14. Quasar Framework. URL: <https://quasar.dev/>
15. Home | Vue Router. URL: <https://router.vuejs.org/>
16. What is Vuex? | Vuex. URL: <https://vuex.vuejs.org/>
17. Web | ParaView. URL: <https://www.paraview.org/web/>
18. ParaViewWeb. URL: <http://kitware.github.io/paraviewweb/>
19. Visualizer. URL: <https://kitware.github.io/visualizer/>

20. Bondarenko A.A., Kononov E.M., Kosolapov O.A., Polyakov S.V., Yakobovskij M.V. Programmnyj kompleks GIMM_NANO // Nauchnyj servis v seti Internet: vse grani parallelizma: Trudy Mezhdunarodnoj superkomp'yuternoj konferencii (23–28 sentyabrya 2013 g., g. Novorossiysk). M.: Izd-vo MGU, 2013. S. 333–337. URL: <http://agora.guru.ru/abrau2013/pdf/333.pdf>

21. Polyakov S.V., Podryga V.O., Puzyr'kov D.V., Kudryashova T.A. Superkomp'yuternoe molekulyarnoe modelirovanie gazodinamicheskogo napyleniya nanostochastic na podlozhku // Superkomp'yuternye dni v Rossii: Trudy mezhdunarodnoj konferencii (24–25 sentyabrya 2018 g., g. Moskva). M.: Izd-vo MGU, 2018. S. 782–792. URL: <http://russianscdays.org/files/pdf18/782.pdf>

22. Podryga V.O., Polyakov S.V. Multiscale mathematical modeling of the metal nanoparticles motion in a gas flow // In: Dimov I., Farago I., Vulkov L. (eds). Seventh International Conference on Finite Difference Methods. Theory and Applications. FDM 2018. Lecture Notes in Computer Science. Cham: Springer, 2019. V. 11386. P. 387–394. https://doi.org/10.1007/978-3-030-11539-5_44

23. Puzyr'kov D., Podryga V., Polyakov S., Yakobovskii M. KIAM_JOB_CONTROL task management environment and its application to cloud and GRID computing) // Selected Papers of the 7th International Conference Distributed Computing and Grid-technologies in Science and Education (GRID 2016), ed. by V. Korenkov, T. Zaikina, F. Nechaevskiy, July 4–9, 2016, Dubna, Russia. CEUR Workshop Proceedings. 2017. V. 1787. P. 416–422. URL: <http://ceur-ws.org/Vol-1787/416-422-paper-72.pdf>

24. Kornilina M.A., Podryga V.O., Polyakov S.V., Puzyr'kov D.V., Yakobovskiy M.V. Cloud service for solution of promising problems of nanotechnology // Supercomputing Frontiers and Innovations. 2017. V. 4, No. 4. P. 66–79. <https://doi.org/10.14529/jsfi170405>

25. Polyakov S.V., Podryga V.O., Kudryashova T.A. HPC simulation of non-linear processes in microsystems gas-metal // Lobachevskii Journal of Mathematics. 2020. V. 41, No. 8. P. 1554–1562. <https://doi.org/10.1134/S1995080220080168>

26. Podryga V.O., Polyakov S.V., Tarasov N.I. Developing of multiscale approach to HPC-simulation of multiphase fluid flows // Lobachevskii Journal of Mathematics. 2021. V. 42, No. 11. P. 2626–2636. <https://doi.org/10.1134/S1995080221110160>

27. Tarasov N., Karamzin Yu., Kudryashova T., Podryga V., Polyakov S. Computer modeling of air flows purification using sorption filters // Journal of Physics: Conf. Series. 2021. V. 2028. Paper 012025 (7 p.).

<https://doi.org/10.1088/1742-6596/2028/1/012025>

28. Kudryashova T.A., Polyakov S.V., Tarasov N.I. Mathematical modelling of electrophysical water treatment // Defect and Diffusion Forum. 2021. V. 412. P. 149–162.

<https://doi.org/10.4028/www.scientific.net/DDF.412.149>

СВЕДЕНИЯ ОБ АВТОРАХ



ТАРАСОВ Никита Игоревич – исследователь, младший научный сотрудник ИПМ им. М.В. Келдыша РАН.

Nikita Igorevich TARASOV – researcher, junior researcher of KIAM RAS.

email: nikita_tarasov@imamod.ru

ORCID 0000-0002-2342-1292



ПОДРЫГА Виктория Олеговна – доктор физико-математических наук, доцент, ведущий научный сотрудник ИПМ им. М.В. Келдыша РАН.

Viktoriia Olegovna PODRYGA – Doctor of Physics and Mathematics, Associate Professor, Leading Researcher of KIAM RAS.

email: pvictoria@list.ru

ORCID 0000-0001-7874-6978

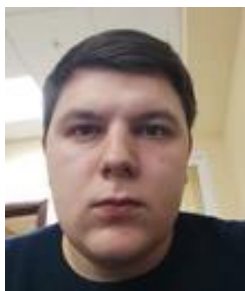


ПОЛЯКОВ Сергей Владимирович – доктор физико-математических наук, старший научный сотрудник, ведущий научный сотрудник ИПМ им. М.В. Келдыша РАН.

Sergey Vladimirovich Polyakov – Doctor of Physics and Mathematics, Associate Professor, Leading Researcher of KIAM RAS.

email: polyakov@imamod.ru

ORCID 0000-0003-1859-9034



ТИМАКОВ Алексей Валерьевич – аспирант МФТИ (ГУ), инженер ИПМ им. М.В. Келдыша РАН

Alexey Valerievich TIMAKOV – Post-graduate student of the Moscow Institute of Physics and Technology (GU), engineer of KIAM RAS

email: aletimakov@yandex.ru

ORCID 0000-0001-6937-8018

Материал поступил в редакцию 15 декабря 2022 года