

ОГЛАВЛЕНИЕ

О. М. Атаева, В. А. Серебряков, Н. П. Тучкова ОБ ОПИСАНИИ НЕКОТОРЫХ КРАЕВЫХ ЗАДАЧ В СЕМАНТИЧЕСКОЙ БИБЛИОТЕКЕ LIVMETA	2–21
С. А. Власова, Н. Е. Каленов, А. Н. Сотников ОНТОЛОГИЯ ВСПОМОГАТЕЛЬНЫХ И ПОЛИТЕМАТИЧЕСКИХ ПРЕДМЕТНЫХ КЛАССОВ ЕДИНОГО ЦИФРОВОГО ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ	22–42
Т. А. Полилова ПЕРЕЧЕНЬ ВАК: ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ В БАЗЕ РНЖ И ELIBRARY.RU	43–64
Ю. Е. Поляк СУБЪЕКТИВНЫЕ ЗАМЕТКИ О ПОИСКОВЫХ СИСТЕМАХ	65–97
И. Н. Соболевская, А. Н. Сотников ВИРТУАЛЬНАЯ ВЫСТАВКА КАК СРЕДСТВО ИНТЕГРАЦИИ В ЕДИНОЕ ЦИФРОВОЕ ПРОСТРАНСТВО НАУЧНЫХ ЗНАНИЙ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ В ОБЛАСТИ НАУКИ И КУЛЬТУРЫ	98–114
Н. П. Тучкова, К. П. Беляев, Г. М. Михайлов СТОХАСТИЧЕСКИЙ АНАЛИЗ ПОТОКОВ ТЕПЛА «ОКЕАН –АТМОСФЕРА» В СЕВЕРНОЙ АТЛАНТИКЕ	115–126
Б. Я. Штейнберг НА ПУТИ К СОЗДАНИЮ РАСПАРАЛЛЕЛИВАЮЩИХ КОМПИЛЯТОРОВ НА ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ С РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ	127–149

УДК 004.822, 517.9

ОБ ОПИСАНИИ НЕКОТОРЫХ КРАЕВЫХ ЗАДАЧ В СЕМАНТИЧЕСКОЙ БИБЛИОТЕКЕ LIBMETA

О. М. Атаева¹ [0000-0003-0367-5575], В. А. Серебряков² [0000-0003-1423-621X],

Н. П. Тучкова³ [0000-0001-5357-9640]

¹⁻³Вычислительный центр им. А.А. Дородницына ФИЦ Информатика и управление РАН, ул. Вавилова, 40, г. Москва, 119333

¹oli.ataeva@gmail.com, ²serebrvas@gmail.com, ³natalia_tuchkova@mail.ru

Аннотация

Использован подход онтологического проектирования для описания семантики некоторых краевых задач в цифровой библиотеке LibMeta.

Для описания задач в библиотеке LibMeta установлены связи терминов и понятий с классическими определениями математической энциклопедии и других первоисточников. Установление связей позволяет сформировать словарь и тезаурус прикладной предметной области краевых задач и поместить результаты в семантическую среду цифровой библиотеки. Примеры такого подхода продемонстрированы с использованием возможностей семантической библиотеки LibMeta, в которую в процессе развития интегрированы в оцифрованном виде версия математической энциклопедии, энциклопедии математической физики, классификаторы, прикладные математические тезаурусы и словари. Новые термины из публикаций после добавления в контент библиотеки отразились со связями в математической энциклопедии. Тезаурус для задач предметной области теории упругости был впервые создан путем интеграции предметных словарей, классификаторов, метаданных публикаций профильного журнала и энциклопедического контента библиотеки LibMeta. Цель таких исследований состоит в предоставлении пользователю дополнительных сервисов в поиске публикаций в прикладной научной области.

Ключевые слова: *прикладная онтология, тезаурус предметной области, источники данных, разработка онтологии, цифровая семантическая библиотека LibMeta*

ВВЕДЕНИЕ

Структурированные данные трактуются в информационных технологиях как знания. В работе мы обращаемся к *оцифрованным научным, математическим знаниям*, которые собраны в математической энциклопедии, классических монографиях и рецензированных публикациях. Работа посвящена вопросам описания в цифровой среде задач математической физики. Онтологическое проектирование позволяет реализовать структурирование этих данных с учетом их смысловой составляющей. Назначение такого проектирования – в создании описания предметной области (ПрО) *для информационного сопровождения научных исследований*, что необходимо в связи с ускорением развития математики и науки в целом.

Задача семантического описания математики решается не впервые, многие отечественные и зарубежные исследователи продвинулись в этом направлении. Не уменьшая значимости других проектов, отметим только наиболее известный проект глобальной математической библиотеки [1]. Надо отметить, что описание задач математической физики – это бесконечная тема в силу обширной сферы их применения, и каждый новый результат составляет очередной шаг в этих исследованиях.

Семантическое описание в работе основано на разработке тезауруса ПрО [2]. Поиск и навигация по ПрО в семантической библиотеке организована с помощью графа знаний [3].

Идея навигации по заранее систематизированной информации существовала, начиная с первых цифровых коллекций, и была развита на основе онтологического проектирования [4–6] и семантического подхода [7], которые логически привели сегодня к навигации по дереву графа.

Технология графа знаний при информационном запросе в семантической библиотеке LibMeta для выбранной ПрО обсуждается в статье на примерах уравнений смешанного типа и уравнений теории упругости.

Специфика современного представления ПрО краевых задач изучалась на основе публикаций специалистов научных журналов, классических монографий, энциклопедии математической физики под ред. академика Л.Д. Фаддеева [8] и математической энциклопедии под ред. академика И.М. Виноградова [9]. В процессе развития библиотеки эти ресурсы в оцифрованном виде, включая

формулы, интегрированы в семантическую библиотеку LibMeta. Специалисты – сотрудники журналов и коллеги – оказывали консультации и высказывали экспертные мнения при работе над описанием выбранных ПрО.

Вопросам представления уравнений смешанного типа в цифровой среде авторы уже посвящали исследования ранее. Подробно эти вопросы были изложены в работах [10–13]. В работе, ссылаясь на эти исследования, мы демонстрируем конкретную их реализацию в библиотеке LibMeta. Были проведены исследования *для уравнений смешанного типа теории упругости* и продолжены исследования *предметной области теории упругости* (ПрО ТУ) для ее интеграции в цифровую среду библиотеки LibMeta. В работе показано, как формируется тезаурус для новой ПрО в LibMeta и строится граф знаний для навигации по ПрО.

Для семантического описания ПрО был сформирован представительный список понятий и соответствующих им терминов, были установлены связи между ними для построения тезауруса и онтологии.

Отличие предлагаемого исследования заключается в *интегрированном подходе* в развитии семантической цифровой персональной библиотеки LibMeta, когда добавление новых разделов математики приводит к насыщению связями онтологии библиотеки, достраиванию графа знаний локальной и/или нескольких ПрО.

1. ОПИСАНИЕ СТАТЕЙ ТЕЗАУРУСА

Проблемы информационного представления математических знаний непосредственно связаны с описанием логических структур, задаваемых внутри математических ПрО. В цифровом представлении логика ПрО задается посредством отношений (семантических связей) в тезаурусе. Тезаурус составляет основу для онтологического проектирования в семантической библиотеке.

Устойчивые словосочетания и структура изложения, характерные для большинства математических научных трудов, известны и предлагаются как стандарт издательствами, на них можно опираться при первичной обработке текстов статей. Для учета специфики ПрО необходимо изучить текстовые конструкции (контексты терминов), характерные именно для конкретной ПрО, на основе представительного списка трудов.

Для понимания логики ПрО краевых задач математической физики обра-

тимся к определению академика В.С. Владимирова [14]: «*Математическая физика – это теория математических моделей физических явлений. Она относится к математическим наукам; критерий истины в ней – математическое доказательство. Однако, в отличие от чисто математических наук, в МФ исследуются физические задачи на математическом уровне, а результаты представляются в виде теорем, графиков, таблиц и т. д. и получают физическую интерпретацию. При таком широком понимании математической физики к ней следует относить и такие разделы механики, как теоретическая механика, гидродинамика и теория упругости*».

Таким образом, при описании ПрО краевых задач математической физики (ПрО УРМФ) надо выделять понятия для *моделей физических явлений, физических задач, математических уравнений*. Эти особенности позволяют перейти на язык онтологий для описания ПрО уравнения смешанного типа (ПрО УЧП СМТ) как части ПрО уравнения с частными производными (ПрО УЧП) и, соответственно, ПрО уравнений математической физики (ПрО УРМФ). То есть можно использовать следующую иерархию для уравнений смешанного типа

$$\text{ПрО УЧП СМТ} \subset \text{ПрО УЧП} \subset \text{ПрО УРМФ}.$$

1.1. Структура статьи тезауруса для уравнений смешанного типа математической физики

Уравнения смешанного типа являются частью предметной области уравнений в частных производных (УЧП) и принимают гиперболический тип, параболический тип или эллиптический тип в зависимости от области определения (рассмотрения). По результатам предыдущих исследований для уравнений смешанного типа в библиотеку был добавлен раздел «*словарь уравнений смешанного типа*» и описан тезаурус [10]. Благодаря наполнению контента библиотеки новыми данными [15] и применению векторных алгоритмов при обработке статей, произошло насыщение онтологии новыми связями. Понятия словаря [10] также получили новые связи. На Рис. 1. показано, что в описании понятия «Трикоми уравнение» появились ссылки на связи, которые возникли при добавлении этих понятий в граф знаний с вершиной УДК (517.956.6).

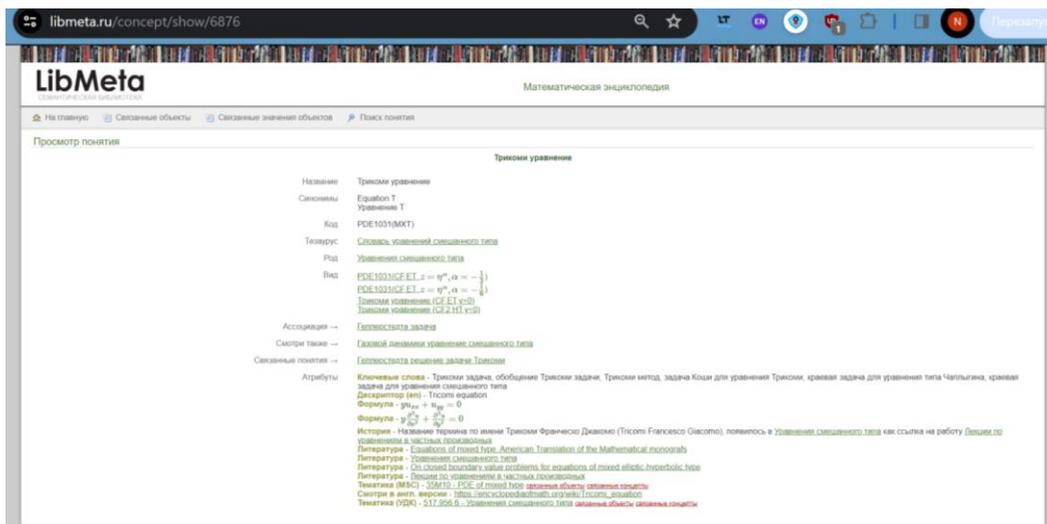


Рис. 1. Просмотр понятия «Трикоми уравнение» со связями в LibMeta

На Рис. 2. показаны объекты, связанные с вершиной «517.956.6 – Уравнения смешанного типа (УДК)».

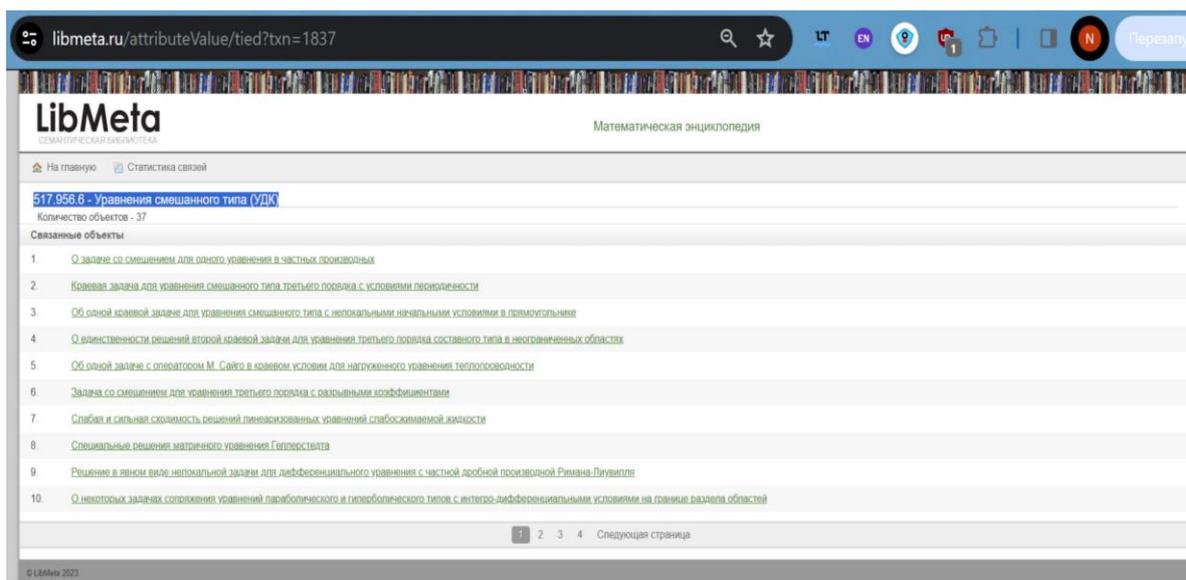


Рис. 2. Просмотр связанных объектов понятия «Трикоми уравнение» в библиотеке LibMeta

Добавление новых публикаций в Про краевых задач позволило также обозначить лакуны в цифровой версии математической энциклопедии. Например, термин «Пенлеве тест» отсутствовал, хотя термин «Пенлеве теорема» был в энциклопедии. Включение метаданных публикации (Рис. 3) привело к расширению математической энциклопедии в LibMeta и добавлению недостающего термина

и соответствующего описания тезауруса LibMeta. Новый термин получил все связи, такие как УДК, связь с англоязычной версией энциклопедии, и т.д., (Рис. 4).

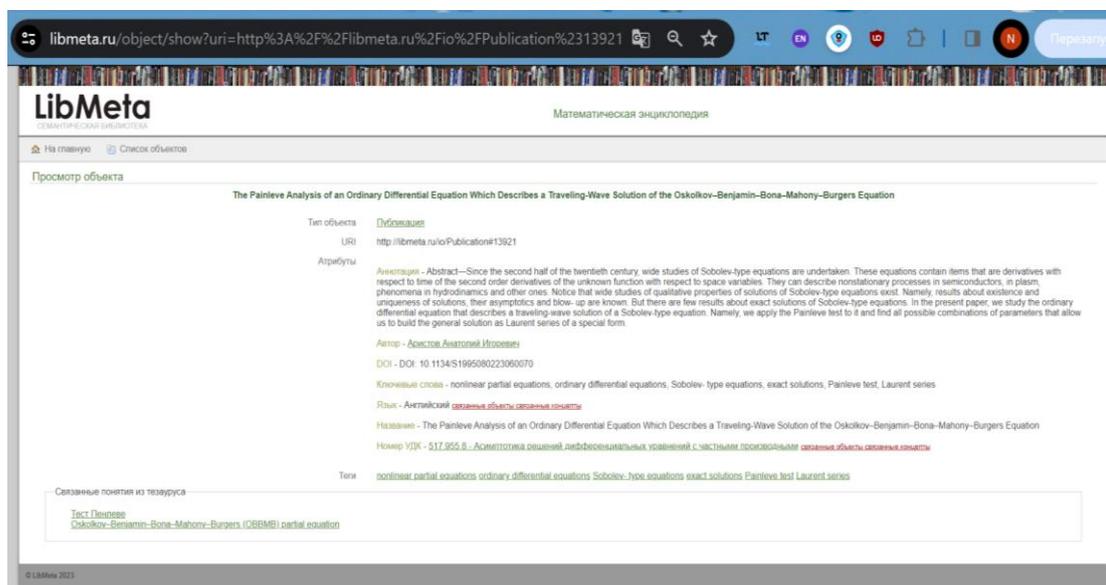


Рис. 3. Просмотр описание добавленной публикации в LibMeta

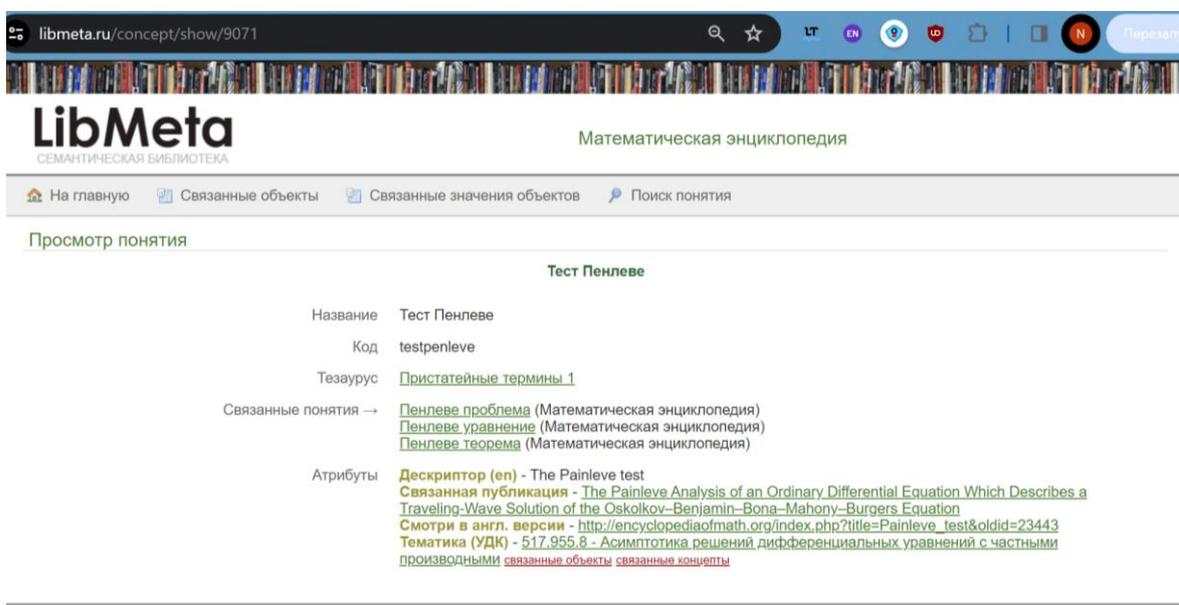


Рис. 4. Просмотр добавленного к математической библиотеке термина

1.2. Структура статьи тезауруса для теории упругости

Специфика современного представительства предметной области *краевых задач теории упругости* в цифровой среде изучалась на основе контента науч-

ного журнала «Механика композиционных материалов и конструкций» (МКМК, <https://mkmk.ras.ru/>).

Для создания описания, кроме энциклопедий [8, 9], были использованы первоисточники классиков этого направления [17–20]. ПрО ТУ – также новая область для библиотеки LibMeta, опишем статьи тезауруса для этой предметной области.

Для термина «задач теории упругости» в качестве основных понятий были выбраны задачи и их видовые формы (частные случаи).

Основные термины – формулировки задач в виде уравнений математической физики (см. Таблицы 1 и 2 как примеры статей тезауруса «Дифференциальные уравнения линейной теории упругости»).

Таблица 1. Статья для термина «Основные уравнения линейной теории упругости. Дифференциальные уравнения линейной теории упругости».

Метка	Описание	Идентификатор
D:	Основные уравнения линейной теории упругости. Дифференциальные уравнения линейной теории упругости.	EqTE000
BT:	Дифференциальные уравнения с частными производными	PDE000
NT:	Уравнения статики в объеме. Уравнения равновесия.	EqTE001
NT:	Определение линейного тензора деформации через вектор перемещения. Геометрическая сторона задачи, связь 3-х компонент перемещений с 6-ю компонентами симметричного тензора деформации.	EqTE002
NT:	Закон состояния линейно-упругого тела. Обобщенный закон Гука для изотропного тела в изотермическом или адиабатическом процессах. Физическая модель теории упругости, связь компонентов тензора напряжения с компонентами тензора деформации.	EqTE003
SeeAlso:	Гипотеза <i>сплошности (сплошной среды)</i> (link MathEnc)	
Ref:	Лурье А.И. Теория упругости. М.: Наука, 1970 (link p.124)	RefTE001

Таблица 2. Статья термина «Краевые задачи линейной теории упругости»

Метка	Описание	Идентификатор
D:	Краевые задачи линейной теории упругости	EqTE200
BT:	Краевые задачи для уравнений математической физики	BdMF001
NT:	Первая задача. Для системы уравнений EqTE001 , EqTE002 , EqTE003 .	EqTE201
NT:	Вторая задача. Для системы уравнений EqTE001 , EqTE002 , EqTE003 .	EqTE202
NT:	Третья задача. Для системы уравнений EqTE001 , EqTE002 , EqTE003 .	EqTE203
Ref:	RefTE001 (link, p.124)	
SeeAlso:	Решения задач теории упругости.	EqTE300
SeeAlso:	Приведение к краевым задачам для уравнений Лапласа и Пуассона. (link MathEnc) RefTE001 (link p. 388)	
SeeAlso:	Задача о кручении. RefTE001 (link p. 388)	
SeeAlso:	Изгиб силой. RefTE001 (link p. 430)	
SeeAlso:	Задача Мичелла. RefTE001 (link p. 443)	

Замечание 1. Погружение в терминологию предметной области означает принятие некоторых нестандартных терминов, установившихся у специалистов, например, термины «сплошность», «условие сплошности». В обычных текстах эти термины могли бы считаться неправильными с точки зрения русского языка, но они характерны для *теории упругости*.

Замечание 2. Для навигации по первоисточникам исследуемой предметной области в цифровой библиотеке создан и пополняется список литературы, который также может быть использован для цитирования читателями. Отметим один из наиболее полных в русскоязычном сегменте список по механике – это (<https://mechmath.ipmnet.ru/lib/?s=solid>), ресурс факультете Мехмат МГУ имени М.В. Ломоносова. Тем не менее, наличие специального списка литературы в цифровой библиотеке LibMeta отличается навигацией по терминам из первоисточников, что можно проследить на примерах Таблиц 1 и 2. Эта особенность позволяет организовать отсылку к конкретному контексту термина в первоисточнике, использовать связь тезауруса **SeeAlso** и связи метаданных источников. Эти свойства списка литературы в LibMeta также позволяют пользователю самому составить мнение о достоверности предъявленных описаний понятий и терминов и/или самостоятельно получить дополнительную информацию в первоисточнике.

В Таблице 3 приведен фрагмент списка литературы по теории упругости, на которую есть ссылки в Таблице 3. Список сформирован в порядке появления ссылок при описании терминов (порядок возрастания номера во второй части идентификатора) и учитывает первую букву фамилии в метке и количество меток с этой буквой в английской транскрипции независимо от языка публикации. Такая логика списка позволяет продлевать его, не переформатируя по алфавиту, и организовывать связи с помощью идентификатора.

Таблица 3. Литература по «теории упругости»

Метка	Библиографическое описание	Идентификатор
L1:	Лурье А.И. Теория упругости. М.: Наука, 1970, 940 с.	RefTE001
L2:	Лурье А.И. Пространственные задачи теории упругости. Гостехиздат, 1955, 491с.	RefTE002
L3:	Лурье А.И. Нелинейная теория упругости. М.: Наука, 1980, 512 с.	RefTE003
T1:	Тимошенко С.П. Курс теории упругости. Киев: Наук. думка, 1972.	RefTE004
M1:	Мейз Дж. Теория и задачи механики сплошных сред. М.: Мир, 1974.	RefTE005
S1:	Седов Л.И. Механика сплошной среды. Т. 1, 2. М.: Наука, 1970	RefTE006
V1:	Vasiliev V.V., Morozov E.V. Advanced Mechanics of Composite Materials and Structures. Fourth Edition, Elsevier, 2018, 856 p.	RefTE100

Таблица 4. Статья термина «Решения задач теории упругости»

Метка	Описание	Идентификатор
D:	Решения задач теории упругости.	EqTE300
NT:	Решение в перемещениях : находится вектор перемещения и по нему вычисляется тензор деформации. Чаще применяется к первой задаче EqTE201	
NT:	Решение в напряжениях : находится такой статически возможный тензор напряжения T , что, определяемый по нему тензор деформации e , удовлетворяет условию <i>сплошности</i> .	

Решения задач теории упругости в самом общем виде различаются постановкой «в перемещениях» или «в напряжениях». Таким образом, термин «Решения задач теории упругости» представляет собой главный дескриптор к остальным формулировкам.

Замечание 3. В Таблицах 1–4 также встречается указание на связь с математической энциклопедией[9] *link MathEnc* для терминов из теории упругости, которые в ней присутствуют.

На Рис. 5 приведен пример понятия из тезауруса для ПрО ТУ.

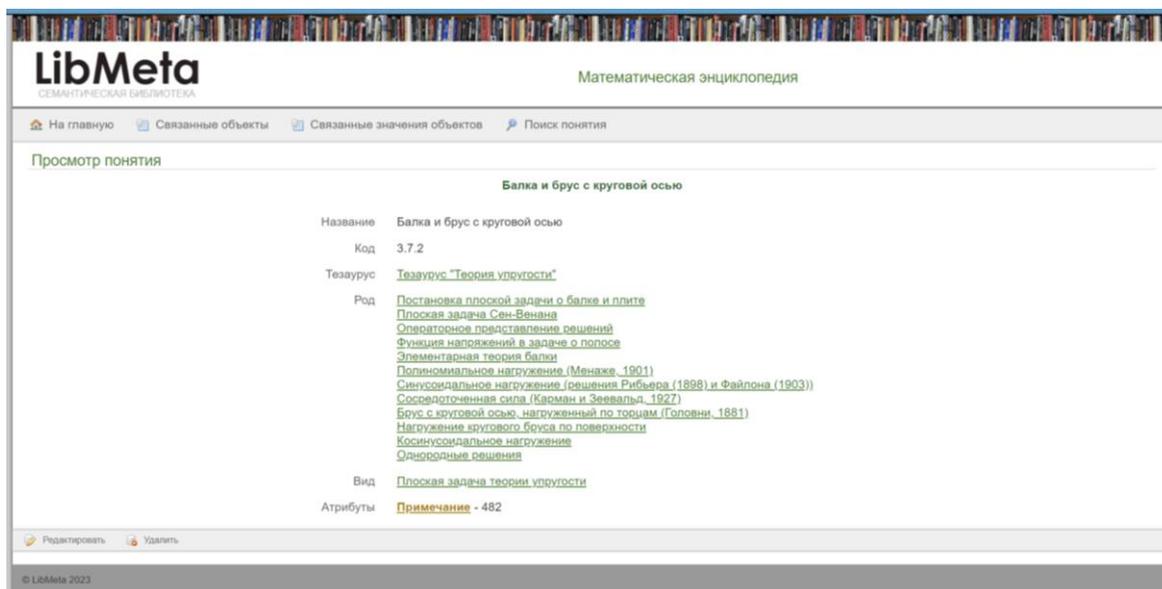


Рис. 5. Понятие из тезауруса ПрО ТУ в LibMeta

В результате, с учетом особенностей предметной области получаем следующий состав статьи тезауруса «задача теории упругости»:

- *идентификатор термина(понятия);*
- *объекты исследования (описание материала, конструкции);*
- *постановка задачи (модель, процесс);*
- *математическая постановка задач (уравнение);*
- *методы решения (численные, аналитические);*
- *решение (представление решения в виде аналитическом, численном, спецфункций и т. д.);*
- *авторы (первоисточников и статей журнала);*
- *ключевые слова (из первоисточников и статей);*
- *синонимы;*
- *видовые термины (частные случаи задач);*
- *ассоциативные термины (все термины из статей, где встречается эта задача);*

- примечания (любые комментарии к статье тезауруса, которые не подходят к перечисленным позициям статьи).

2. РЕАЛИЗАЦИЯ СЕМАНТИЧЕСКОГО ПОДХОДА В БИБЛИОТЕКЕ LIBMETA

Были обработаны текстовые документы, которые содержали информацию о ГОСТе (<https://docs.cntd.ru/document/1200113813>), предметном указателе [16] и его оглавлении, а также список задач теории упругости и связанные с ними решения, уравнения, методы, список дескрипторов и список связей.

В качестве внешних базовых таксономий, с которыми устанавливаются связи, использованы математическая энциклопедия, энциклопедия математической физики, тезаурус обыкновенных дифференциальных уравнений (ОДУ), словарь специальных функций математической физики и другие источники и составляющие библиотеки.

Для размещения в библиотеку LibMeta тексты статей подвергались предварительной обработке. Решались задачи по очистке и предварительной обработке, извлечения именованных сущностей и т. д. [21]. Предварительно при необходимости выполнялось приведение слов к именительному падежу, единственному числу, начальной форме глагола, мужскому роду. Затем проводится сравнение терминов, при этом порядок слов не учитывается. Если совпадение полное, то устанавливалась связь типа «*sameAs*». Если совпадение терминов не полное, то выполнялся поиск понятия, которое соответствует совпавшим словам. Если такой термин есть, то устанавливается связь «*seeAlso*» и «*narrower*». В результате на основе онтологии были получены файлы в формате *xml*, пригодные для пакетной загрузки в систему.

2.1 ФОРМИРОВАНИЕ ТЕЗАУРУСА ДЛЯ ЗАДАЧ ТЕОРИИ УПРУГОСТИ

Термины в тезаурусе для задач теории упругости были расставлены в соответствии с разделами: «Задача», «Метод», «Решение», «Материал», «Уравнение», поскольку они составляют особенности статей в ПрО МКМК. Часть работы по составлению разделов была выполнена вручную, часть связей сформировалась автоматически на основе связей между терминами тезауруса. Если коды классификаторов были проставлены у «родовых» понятий, то термины, которые связались по связям «*sameAs*», получают код автоматически. Работа по формированию этих разделов является итеративным процессом и зависит от новых по-

ступающих публикаций журнала.

2.2 ПРИМЕР ДОБАВЛЕНИЯ СТАТЬИ В ТЕЗАУРУС

В математической энциклопедии встречается понятие «*Краевая задача*», которое является *родовым* для следующих понятий из математической энциклопедии:

1. внешняя и внутренняя краевые задачи,
2. первая краевая задача,
3. аппроксимация дифференциальной краевой задачи разностной,
4. коэрцитивная краевая задача,
5. нелинейная краевая задача,
6. линейная краевая задача,
7. третья краевая задача,
8. вторая краевая задача,
9. смешанная и краевая задачи для гиперболических уравнений и систем,
10. смешанная и краевая задачи для параболических уравнений и систем.

В результате связывания с понятиями теории упругости у понятия математической энциклопедии «*Краевая задача*» появляются связи типа «*seeAlso*» со следующими понятиями из тезауруса ПрО МКМК:

1. Интегральные уравнения первой краевой задачи;
2. Интегральные уравнения второй краевой задачи;
3. Сопоставление интегральных уравнений первой и второй краевых задач;
4. Вторая внутренняя краевая задача;
5. Первая внешняя краевая задача;
6. Краевая задача для двусвязного тела вращения;
7. Первая краевая задача для полупространства;
8. Первая краевая задача;
9. Вторая краевая задача для сферы;
- 10....

При этом каждое из этих понятий ПрО МКМК является *видовым* понятием для рассматриваемого понятия из математической энциклопедии. Каждое из них в свою очередь может образовывать свои связи с математической энциклопедией. Например, понятие из математической энциклопедии «*Интегральное*

уравнение» оказывается связанным с понятием из тезауруса ПрО МКМК «Интегральные уравнения первой краевой задачи» и одновременно оказывается для него *родовым* понятием. Связи «*seeAlso*» устанавливаются изначально между понятиями разных источников. На Рис. 6 представлен фрагмент тезауруса с установленными связями между понятиями из различных источников.

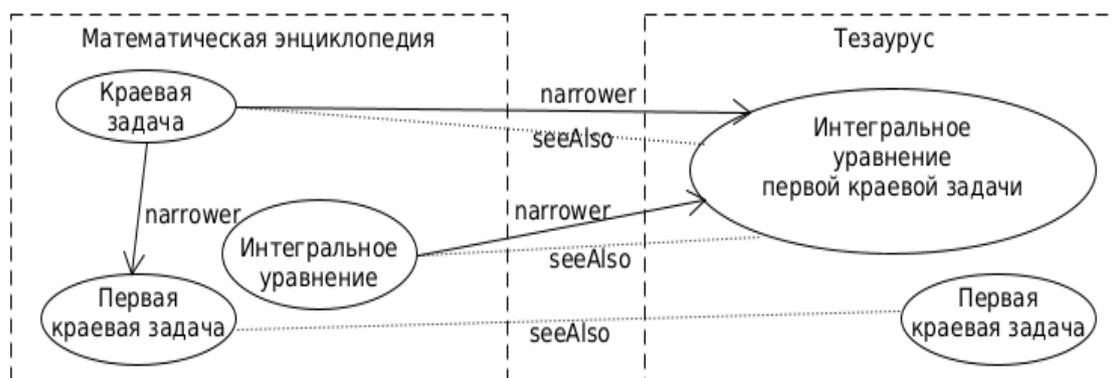


Рис. 6. Фрагмент тезауруса ПрО МКМК со связями между понятиями из различных источников

При дополнительном анализе эти понятия могут связываться дополнительными связями, как в случае, представленном на Рис. 7, новая связь «*narrower*» (видовое понятие) демонстрирует *иерархическую* связь между понятиями.

Включение задач теории упругости в библиотеку LibMeta расширило возможности навигации по контенту библиотеки. На примере странички публикации из МКМК (Рис. 8), можно видеть, что пользователю предоставляются дополнительные поля для навигации по терминам. Показано, что появились связи с различными первоисточниками, которых раньше в библиотеке не было.

3. НАВИГАЦИЯ ПО КОНТЕНТУ БИБЛИОТЕКИ

Представление предметной области в виде тезауруса (онтологии) составляет основу для формирования графа знаний для организации навигации по данным библиотеки. Для размещения нового объекта в граф происходит предобработка данных, которая заключается в формировании метаданных и подготовке текстов к включению в инфраструктуру графа знаний. В качестве объекта могут выступать разные типы узлов графа знаний такие, как публикации, авторы, формулы, теоремы и т. д. Классификаторы могут дополняться (углублять) эле-

ментами из разных (других, например, УДК дополняется MSC) классификаторов. В качестве понятий на концах ребер графа могут также выступать понятия разных тезаурусов (частей тезауруса, «кустов» тезауруса). На Рис. 7 представлена схема такого графа, где сплошные линии представляют связи публикации с контентом библиотеки, а пунктирные – заданные в онтологии библиотеки.

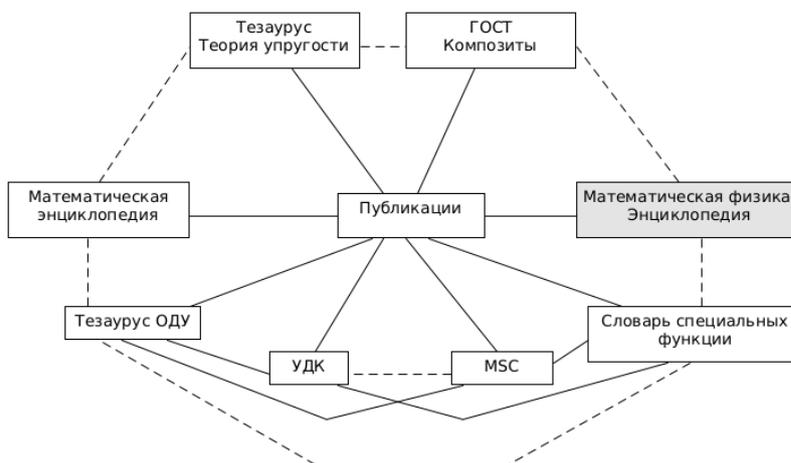


Рис. 7. Схема графа знаний для узла «публикация»

Каждому элементу данных, загруженных в библиотеку, можно сопоставить вершину онтологии, определяющую положение элемента данных в онтологии. Для навигации по графу знаний предметной области журнала МКМК естественно выбрать «публикации» в качестве узлов графа.

Реальное изображение публикации в библиотеке LibMeta и связей для навигации показано на Рис. 8. Центральное изображение – это страничка публикации, а левый нижний столбец – список терминов, встречающихся в тексте статьи, которые можно применять для навигации по терминам из тезауруса. Список сформирован при обращении к публикации, поскольку она связана с тезаурусом Про ТУ. Навигация по терминам также обеспечивается пересечением метаданных в LibMeta, а не только связями в тезаурусе.

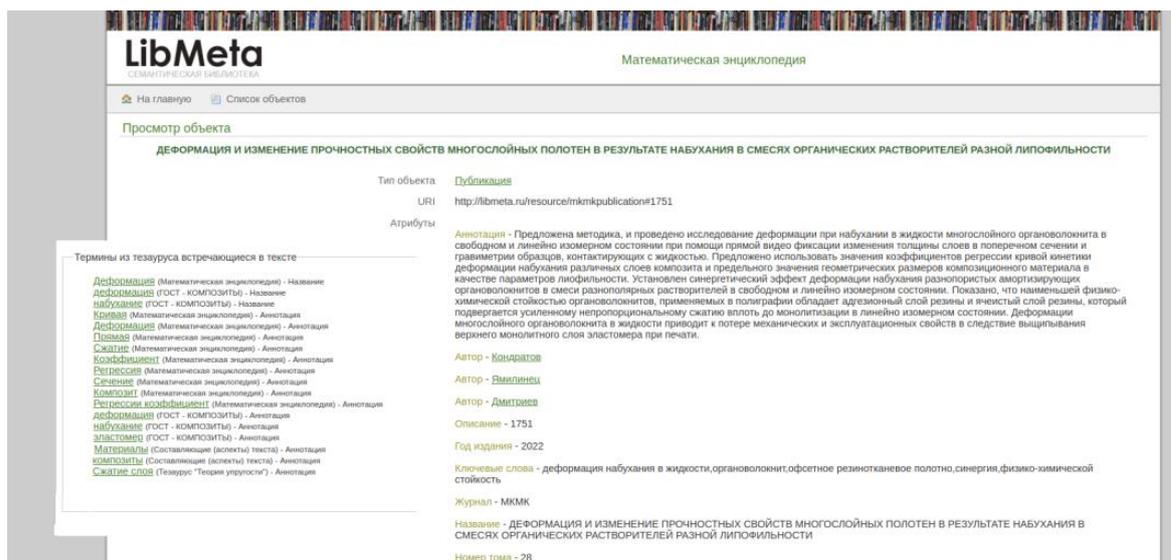


Рис. 8. Пример странички публикации из МКМК с терминами для навигации

ЗАКЛЮЧЕНИЕ

Предложенная работа содержит некоторые результаты исследований вопросов представительства русскоязычного научного контента в цифровых информационных ресурсах. Дальнейшие исследования направлены на персонафикацию предметных областей для предоставления пользователю возможностей навигации по авторским публикациям с учетом специфики информационного запроса, таких как задачи, методы, решения и т. д.

Полученные результаты могут быть использованы в численном и аналитическом моделировании задач математической физики. Возможность навигации по знаниям в цифровой среде может улучшить информационное сопровождение научных исследований. Для этого цифровая библиотека расширяется за счет новых результатов и новых предметных областей.

Часть обработки трудов журнала МКМК была выполнена с применением нейросети на вычислительном комплексе ФИЦ «Информатика и управление» РАН ЦКП «Информатика».

Работа представлена в рамках выполнения темы НИР 0063-2019-0003 ФИЦ Информатика и управление РАН.

СПИСОК ЛИТЕРАТУРЫ

1. *Ion P.D.F., Watt S.M.* The Global Digital Mathematics Library and the International Mathematical Knowledge Trust // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds.). Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science. 2017. V. 10383. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-62075-6_5.
2. *Серебряков В.А., Атаева О.М.* Основные понятия формальной модели семантических библиотек и формализация процессов интеграции в ней // Программные продукты и системы. 2015. № 4. С. 180–187.
3. *Liu L., Omidvar A., Ma Z., Agrawal A., An A.* Unsupervised Knowledge Graph Generation Using Semantic Similarity Matching // Proceedings of the Third Workshop on Deep Learning for Low Resource Natural Language Processing. 2022. P. 169–179. <https://doi.org/10.18653/v1/2022.deeplo-1.18>
4. *Gruber T.* Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). 2007. V. 3. No. 1. P. 1–11.
5. *Vrandečić D.* Ontology Evaluation, In Handbook on Ontologies, International Handbooks on Information Systems, edited by S. Staab, R. Studer, 2009. P. 293–313. https://doi.org/10.1007/978-3-540-92673-3_13.
6. *Елизаров А.М., Кириллович А.В., Липачёв Е.К., Невзорова О.А.* Онтология математического знания OntoMathPRO // Доклады Российской академии наук. Математика, информатика, процессы управления. 2022. Т. 507. № 1. С. 29–35. <https://doi.org/10.31857/S2686954322700011>
7. Web. <https://www.w3.org/standards/semanticweb>.
8. Математическая физика. Энциклопедия. Гл. ред. Л.Д. Фаддеев. М.: Большая Российская энциклопедия, 1998.
9. *Виноградов И.М.* (ред.) Математическая энциклопедия (в 5 томах). М.: Советская энциклопедия. 1977–1985.
10. *Моисеев Е.И., Муромский А.А., Тучкова Н.П.* О тезаурусе предметной области смешанные уравнения математической физики // CEUR Workshop Proceedings. 2018. V. 2260. P. 395–405.

11. Муромский А.А., Тучкова Н.П. Представление математических понятий в онтологии научных знаний // Онтология проектирования. 2019. Т. 9. № 1 (31). С. 50–69. <https://doi.org/0.18287/2223-9537-2019-9-1-50-69>.
12. Атаева О.М., Серебряков В.А., Тучкова Н.П. Подходы к организации математических знаний при формировании предметных тезаурусов различных разделов математики // CEUR Workshop Proceedings. 2018. V. 2260. P. 42–54.
13. Ataeva O., Serebryakov V., Tuchkova N. Development of the semantic space 'Mathematics' by integrating a subspace of its applied area // Lobachevskii J. of Mathematics, 2022. V. 43, No. 12. P. 3435–3446. <https://doi.org/10.1134/S1995080222150069>.
14. Владимиров В.С. Что такое математическая физика? Препринт МИАН № НС-06-001, 2006.
15. Ataeva O., Serebryakov V., Tuchkova N. Creating the Applied Subject Area Ontology by Means of the Content of the Digital Semantic Library // Lobachevskii J. of Mathematics, 2022. V. 43, No. 7. P. 1795–1804. <https://doi.org/10.1134/S1995080222100043>
16. Лурье А.И. Теория упругости. М.: Наука, 1970.
17. Лурье А.И. Нелинейная теория упругости. М.: Наука, 1980.
18. Лурье А.И. Пространственные задачи теории упругости. М.: Гостехиздат, 1955.
19. Мейз Дж. Теория и задачи механики сплошных сред. М.: Мир, 1974.
20. Vasiliev V.V., Morozov E.V. Advanced Mechanics of Composite Materials and Structures. Fourth Edition, Elsevier, 2018.
21. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. Автоматическая обработка текстов на естественном языке и анализ данных. М.: ВШЭ, 2017.

ABOUT DESCRIPTIONS OF SOME BOUNDARY-VALUE PROBLEMS IN THE SEMANTIC LIBRARY LIBMETA

O. M. Ataeva¹ [0000-0003-0367-5575], V. A. Serebryakov² [0000-0003-1423-621X],

N. P. Tuchkova³ [0000-0001-5357-9640]

¹⁻³FRS «Computer Sciences and Control», Russian Academy of Sciences, Vavilov str.,
40, Moscow, 119333, Russia

¹oli.ataeva@gmail.com, ²serebrvas@gmail.com, ³natalia_tuchkova@mail.ru

Abstract

An ontological design approach was used to describe the semantics of some boundary value problems in the LibMeta digital library. To describe problems in the LibMeta library, connections of terms and concepts with classical definitions of the mathematical encyclopedia and other primary sources have been established. Establishing links allows you to form a dictionary and thesaurus of the applied subject area of new boundary value problems and place the results in the semantic environment of the digital library. Examples of this approach are demonstrated using the capabilities of the LibMeta semantic library, which contains a digitized version of the mathematical encyclopedia, encyclopedia of mathematical physics, classifiers, and applied mathematical thesauri and dictionaries. New terms from publications, after being added to the content of the library, were reflected with links in the mathematical encyclopedia. The thesaurus for problems in the elasticity theory domain was created for the first time by integrating subject dictionaries, classifiers, metadata of specialized journal publications and encyclopedic content of the LibMeta library. The purpose of such research is to provide the user with additional services in the search for publications in the applied scientific field.

Keywords: *applied ontology, subject domain thesaurus, data sources, ontology development, LibMeta semantic digital library*

REFERENCES

1. *Ion P.D.F., Watt S.M.* The Global Digital Mathematics Library and the International Mathematical Knowledge Trust // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds.). Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science. 2017. V. 10383. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-62075-6_5.
2. *Serebryakov V., Ataeva O.* Osnovnye ponyatiya formal'noj modeli semanticheskikh bibliotek i formalizaciya processov integracii v nej // Software and systems. 2015. № 4. S. 180–187.
3. *Liu L., Omidvar A., Ma Z., Agrawal A., An A.* Unsupervised Knowledge Graph Generation Using Semantic Similarity Matching // Proceedings of the Third Workshop on Deep Learning for Low Resource Natural Language Processing. 2022. P. 169–179. <https://doi.org/10.18653/v1/2022.deeplo-1.18>.
4. *Gruber T.* Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). 2007. V. 3. No. 1. P. 1–11.
5. *Vrandečić D.* Ontology Evaluation, In Handbook on Ontologies, International Handbooks on Information Systems, edited by S. Staab, R. Studer, 2009. P. 293–313. https://doi.org/10.1007/978-3-540-92673-3_13.
6. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* OntoMath-PRO – ontology of mathematical knowledge// Doklady Mathematics. 2022. T. 507. No. 1. S. 29–35. <https://doi.org/10.31857/S2686954322700011>.
7. Semantic Web. <https://www.w3.org/standards/semanticweb>.
8. *Matematicheskaya fizika.* Enciklopediya. Gl. red. L.D. Faddeev. M.: Bol'shaya Rossijskaya enciklopediya, 1998.
9. *Vinogradov I.M.* (red.). Matematicheskaya enciklopediya (v 5 tomah). M.: Sovetskaya enciklopediya. 1977–1985.
10. *Moiseev E.I., Muromskij A.A., Tuchkova N.P.* O tezauruse predmetnoj oblasti smeshannye uravneniya matematicheskoy fiziki // CEUR Workshop Proceedings. 2018. V. 2260. P. 395–405.

11. *Muromskij A.A., Tuchkova N.P.* Predstavlenie matematicheskikh ponyatij v ontologii nauchnyh znaniy // *Ontology of designing*. 2019. T. 9. № 1 (31). C. 50–69. <https://doi.org/0.18287/2223-9537-2019-9-1-50-69>.
12. *Ataeva O., Serebryakov V., Tuchkova N.* Podhody k organizacii matematicheskikh znaniy pri formirovaniya predmetnyh tezaurusov razlichnyh razdelov matematiki // *CEUR Workshop Proceedings*. 2018. V. 2260. P. 42–54.
13. *Vladimirov V.S.* Chto takoe matematicheskaya fizika? Preprint MIAN № NS-06-001, 2006.
14. *Ataeva O., Serebryakov V., Tuchkova N.* Development of the semantic space 'Mathematics' by integrating a subspace of its applied area // *Lobachevskii J. of Mathematics*, 2022. V. 43, No. 12. P. 3435–3446. <https://doi.org/10.1134/S1995080222150069>.
15. *Ataeva O., Serebryakov V., Tuchkova N.* Creating the Applied Subject Area Ontology by Means of the Content of the Digital Semantic Library // *Lobachevskii J. of Mathematics*, 2022. V. 43, No. 7. P. 1795–1804. <https://doi.org/10.1134/S1995080222100043>.
16. *Lurie A.I.* Teoriya uprugosti. M.: Nauka, 1970.
17. *Lurie A.I.* Nelinejnaya teoriya uprugosti. M.: Nauka, 1980.
18. *Lurie A.I.* Prostranstvennye zadachi teorii uprugosti. M.: Gostekhizdat, 1955.
19. *Mais J.* Teoriya i zadachi mekhaniki sploshnyh sred. M.: Mir, 1974.
20. *Vasiliev V.V., Morozov E.V.* Advanced Mechanics of Composite Materials and Structures. Fourth Edition, Elsevier, 2018.
21. *Bolshakova E.I., Vorontsov K.V., Efremova N.E., Klyshinsky E.S., Lukashevich N.V., Sapin A.S.* Automatic natural language text processing and data analysis. M.: HSE, 2017.

СВЕДЕНИЯ ОБ АВТОРАХ



АТАЕВА Ольга Муратовна – старший научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат техн. наук, специалист в области системного программирования и баз данных.

Olga Muratovna ATAeva – senior researcher of the of Dorodnicyn computing center FRC SCS RAS, PhD, expert in the field of system programming and databases.

email: oli.ataeva@gmail.com

ORCID: 0000-0003-0367-5575



СЕРЕБРЯКОВ Владимир Алексеевич – специалист в области теории формальных языков и её приложений, доктор физ.-мат. наук, профессор, зав. отделом Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН. Руководитель и участник разработки ряда известных программных проектов, в частности, ИСИР и ИСИР РАН, «Научный портал РАН».

Vladimir Alekseevich SEREBRYAKOV – expert in the field of theory of formal languages and its applications, doctor of sciences, professor, head of Dorodnicyn computing center FRC SCS RAS department. Head and participant in the development of a number of well-known program projects, in particular, ISIR and ISIR RAS, Scientific portal RAS.

email: serebrvas@gmail.com

ORCID: 0000-0003-1423-621X



ТУЧКОВА Наталия Павловна – старший научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук, окончила ВМиК МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

Natalia Pavlovna TUCHKOVA – senior researcher of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree, graduated from CS Faculty of Lomonosov MSU. The expert in the field of algorithmic languages and information technologies.

email: natalia_tuchkova@mail.ru

ORCID: 0000-0001-5357-9640

Материал поступил в редакцию 26 декабря 2023 года.

УДК 013, 004.65

ОНТОЛОГИЯ ВСПОМОГАТЕЛЬНЫХ И ПОЛИТЕМАТИЧЕСКИХ ПРЕДМЕТНЫХ КЛАССОВ ЕДИНОГО ЦИФРОВОГО ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ

С. А. Власова¹ [0000-0003-1533-5850], Н. Е. Каленов² [0000-0001-5269-0988],

А. Н. Сотников³ [0000-0002-0137-1255]

¹⁻³Межведомственный суперкомпьютерный центр (МСЦ) РАН – филиал ФГУ
ФНЦ Научно-исследовательский институт системных исследований (НИИСИ)
РАН

¹vas.svetlana2013@yandex.ru, ²nekalenov@mail.ru, ³asotnikov@jssc.ru

Аннотация

Одними из основных компонентов Единого Цифрового Пространства Научных Знаний (ЕЦПНЗ) являются предметные онтологии отдельных тематических подпространств, включающие в себя основные понятия, относящиеся к данному научному направлению. Задача построения предметных онтологий на первом этапе требует формирования массива ключевых терминов в заданной области науки с последующим установлением связей между ними. Настоящая работа является развитием исследований, проводимых авторами в области создания ЕЦПНЗ. В рамках предыдущих исследований была предложена унифицированная структура представления онтологии элементов ЕЦПНЗ (подпространств, классов и атрибутов объектов, связей между объектами или атрибутами). В процессе моделирования онтологии на примере универсального и ряда тематических подпространств ЕЦПНЗ выявилась необходимость некоторой корректировки структуры онтологии, касающейся справочников ЕЦПНЗ, для обеспечения возможности описания вложенных атрибутов данных. Кроме того, в онтологию введено понятие «тип словаря значений атрибутов данных», определены два типа словарей – «статические» и «динамические». Эта информация позволяет упростить алгоритмы формально-логического контроля при формировании контента ЕЦПНЗ. Указание

на тип словаря введено в структуру справочников атрибутов объектов. В представленной работе описана модифицированная структура онтологии на примере 11-ти вспомогательных и 10-ти предметных классов универсального подпространства (УПП) ЕЦПНЗ. Приведены примеры справочников каждого класса, построенные в соответствии с моделью структуры онтологии, перечень атрибутов объектов и примеры статических словарей.

Ключевые слова: *цифровое пространство научных знаний, онтология, классы объектов, атрибуты, структуризация, связанные данные.*

ВВЕДЕНИЕ

Настоящая работа является продолжением исследований, проводимых в МСЦ РАН и связанных с созданием Единого цифрового пространства научных знаний (ЕЦПНЗ) как структурированной интегрированной информационной среды, отражающей достижения в различных областях науки [1, 2]. На предыдущих этапах работы была определена архитектура построения ЕЦПНЗ [3], рассмотрены вопросы сетевого обеспечения ЕЦПНЗ [4] и отражения в нем 3D-моделей мультимедийных объектов [5, 6].

В материалах конференции «Научный сервис в сети Интернет» в 2022 году опубликован доклад, описывающий предложенную нами модель структуры онтологии ЕЦПНЗ [7]. Модифицированная версия модели представлена в [8, 9]. Согласно этой модели ЕЦПНЗ представляется в виде 5-ти уровневой иерархической структуры (ЕЦПНЗ – подпространства – классы объектов – атрибуты объектов класса – значения атрибутов), дополненной связями трех типов (универсальные, квазиуниверсальные и специфические). Информация обо всех элементах ЕЦПНЗ хранится в справочниках и словарях. Элементы каждого иерархического уровня описаны справочниками фиксированной для каждого уровня структуры. Справочниками представлены также связи каждого из трех типов. Значения атрибутов и связей хранятся в словарях, информация о которых содержится в соответствующих справочниках. Словари подразделяются на две группы – статические и динамические. Первые содержат значения «стандартизованных» атрибутов (таких как перечень ученых степеней, званий и должностей, рубрик классификационных систем и т. п.). Такие словари заполняются при первоначальной инсталляции ЕЦПНЗ

или при формировании его подпространств и корректируются администраторами, наделенными соответствующими правами.

Словари второго типа наполняются по мере формирования контента ЕЦПНЗ конкретными данными (фамилии персон, наименования публикаций, ссылки на сетевые ресурсы и т. п.).

КЛАССЫ УНИВЕРСАЛЬНОГО ПОДПРОСТРАНСТВА И ИХ АТТРИБУТЫ

Реализация предложенной структуры моделировалась на примере формирования элементов универсального подпространства (УПП) ЕЦПНЗ, содержащего вспомогательные классы объектов, а также объекты мультидисциплинарного характера и связи между ними. В рамках модели выделено 11 вспомогательных и 10 предметных классов объектов.

В качестве вспомогательных выделены следующие классы объектов:

- Форматы данных;
- Универсальные классификационные системы;
- Местоположение (географические характеристики);
- Временные характеристики, даты;
- Количественные характеристики;
- Мировые константы;
- Единицы измерения;
- Языки;
- Группы персон;
- Коллекции;
- Числовые значения.

К предметным классам УПП отнесены следующие:

- Персоны;
- Публикации;
- Квалификационные работы;
- Документы;
- Музейные предметы;
- Изображения и мультимедийные объекты;
- События / мероприятия;
- Организации;

- Политематические базы данных, каталоги ресурсов;
- Награды, гранты.

Для каждого из предметных и вспомогательных классов составлены перечни атрибутов. При определении набора атрибутов объектов учитывался многолетний положительный опыт эксплуатации электронной библиотеки «Научное наследие России» [10], современная версия которой построена на совокупности связанных разнородных данных и развивается как модель составляющей ЕЦПНЗ [11, 12].

Ниже приведен перечень атрибутов объектов УПП. В скобках после наименования атрибута указано, является ли его значение обязательным (о) или факультативным (ф).

Атрибуты вспомогательных классов

Важнейшую роль в структуре онтологии ЕЦПНЗ играет класс «Форматы данных». Его объекты представляют собой структурированный набор правил описания атрибутов других классов. Указание на формат представления данных присутствует во всех справочниках атрибутов объектов и связей. Информация о форматах используется для формально-логического контроля вводимых данных при потоковой загрузке, а также в процессе диалога с оператором «ручного» ввода данных. Кроме того, этот класс объектов может использоваться в качестве источника информации о тех или иных форматах данных, с которыми встречаются пользователи ЕЦПНЗ.

Атрибуты объектов класса «Форматы представления данных»:

- тип представления данных (о); атрибут может принимать значения «текст», «целое число», «дата в формате гггг[.мм[.дд]]», «связи» и т. д.);
- вид формата (ф); атрибут конкретизирует формат, например, «pdf», «jpeg», «URL», ряд значений его словаря описывает структуру представления различных видов связей ЕЦПНЗ;
- обязательное (r) или факультативное (f) значение атрибута (о);
- уникальное (u) или множественное (m) значение атрибута (о);
- ограничения по структуре (ф) (атрибут содержит наименование конкретной структуры, которой должен соответствовать тот или иной атри-

бут, например, «ГОСТ 7-1.2003: Библиографическое описание¹» или «алгоритм контроля ISBN²», или «необходимые требования к структуре адреса электронной почты³» и т. п.);

- ссылка на подробное описание формата (ф).

Атрибуты объектов класса «Универсальные классификационные системы»:

- наименование классификации (о);
- наименование раздела (о);
- индекс (о).

Между объектами этого класса могут устанавливаться различные простые связи универсального типа.

- Атрибуты объектов класса «Местоположение»:
- название объекта (о);
- ид объекта (о);
- дополнительная информация (ф).

Объекты этого класса могут быть связаны с объектами подпространства «География», перечень атрибутов которых существенно шире.

Атрибуты объектов класса «Временные характеристики, даты»:

- название характеристики (о);
- словесное значение характеристики или URN числового значения (ф).

Атрибуты объектов класса «Единицы измерения»:

- наименование единицы измерения (о);
- предмет измерения (о);
- обозначения (аббревиатура) (о);
- дополнительная информация (ф).

Атрибуты объектов класса «Языки»:

- наименование языка (о);
- двухсимвольный код языка (в соответствии со стандартом группы ISO 639) (о);

¹ со ссылкой на официальное описание ГОСТа

² с приведением формулы контроля

³ с формулировкой требований

- трехсимвольный код языка (в соответствии со стандартом группы ISO 639) (о).

Атрибуты объектов класса «Количественные характеристики»:

- наименование характеристики (о);
- единица измерения (о);
- числовое значение (о).
- Атрибуты объектов класса «Мировые константы»:
- наименование (о);
- описание (о);
- обозначение (о);
- значение (о).

Атрибуты объектов класса «Группы персон»:

- наименование группы (о);
- описание (ф).

Атрибуты объектов класса «Коллекции»:

- наименование коллекции (о);
- классы включаемых объектов (о);
- условия включения объекта в коллекцию (о).

Атрибутами объектов вспомогательного класса «Числовые значения» являются любые числа.

Атрибуты предметных классов УПП

- Атрибуты объектов класса «Персоны»:
- фамилия (о);
- имя (о);
- отчество (ф);
- псевдоним (ф);
- дата рождения точная (ф);
- дата рождения приблизительная (ф);
- место рождения (о);
- дата смерти (ф);
- место смерти (ф);
- ученая степень (ф);

- ученое звание (ф);
- биография (о);
- библиография персоны как автора (ф);
- библиография о персоне (ф).

Дополнительная информация (область научных интересов, трудовая деятельность, научные открытия, идентификаторы в базах данных и т. д.) оформляется как именованные связи с соответствующими объектами.

Атрибуты объектов класса «Публикации»:

- наименование (о);
- год издания (о);
- вид (о);
- библиографическое описание (о);
- ISBN (ф);
- ISSN (ф);
- аннотация (ф);
- URL полного текста (ф);
- полный текст (если он представлен в ЕЦПНЗ) (ф).

Дополнительная информация оформляется в виде связей с персонами (может принимать значения «автор», «переводчик», «редактор», «о персоне» и т. д.), организациями, классификационными системами и другими классами объектов.

Атрибуты объектов класса «Квалификационные работы»:

- наименование (о);
- вид работы (о);
- дата выпуска (защиты) (ф);
- URL полного текста (ф);
- Полный текст, если он включен в ЕЦПНЗ (ф).

Дополнительная информация оформляется в виде связей с персонами (автор, научный руководитель, оппонент и т. д.), организациями (место выполнения, место защиты, ведущая организация и т. п.), классификационными системами и пр.

Атрибуты объектов класса «Музейные предметы»:

- наименование объекта (о);
- наименование музея, где хранится оригинал (именованная связь с объектом класса «организации») (о);
- вид источника поступления (о);
- дата поступления в музей (о);
- дата обнаружения (создания) предмета (о);
- описание предмета (о).

Музейные объекты могут быть связаны специфическими связями с персонай (изготовитель объекта, владелец, «автор сбора» – для естественнонаучных коллекций), географическим объектом (место сбора), публикациями и документами, касающимися данного объекта, и т. д.

Атрибуты объектов класса «События / мероприятия»:

- наименование (о);
- тип события (о);
- дата начала события (мероприятия) (о);
- дата окончания события (мероприятия) (о),
- описание события (мероприятия) (ф).

Эти данные дополняются связями с объектами других классов по аналогии с предыдущими. В частности, связь события с персонай (если речь идет о конференции) может принимать значения «председатель оргкомитета», «член программного комитета», «спонсор» и т. п.

Атрибуты объектов класса «Документы»:

- наименование (о);
- вид документа; (о);
- текст документа (ф).

Атрибуты объектов класса «Изображения и мультимедийные материалы»:

- наименование (о);
- вид объекта (о);
- адрес объекта (о);
- дата создания объекта (ф);
- развернутая информация об объекте (ф).

Атрибуты объектов класса «Организации»:

- наименование (о);
- тип организации – научная, образовательная, производственная и т.п. (о);
- вид организации – бюджетная, ООО, ЗАО и т. п. (о).;
- местонахождение (о);
- контактные данные (ф);
- дата создания (ф);
- дата ликвидации (ф).

Атрибуты объектов класса «Политематические базы данных, каталоги ресурсов»:

- наименование (о);
- вид объекта (о);
- условия доступа (ф);
- URL ресурса (о).

Атрибуты объектов класса «Награды, гранты»:

- наименование (о);
- вид награды (о);
- раздел науки (о);
- дата награждения (получения гранта) (ф);
- дата окончания гранта (ф).

Для формирования справочников атрибутов и словарей их значений разработана диалоговая программа, с помощью которой сформированы справочники вышеперечисленных классов объектов и атрибутов, а также элементы статических словарей их значений.

Ниже приведен ряд справочников классов, фрагментов справочников атрибутов и статических словарей.

Class.1: Форматы представления данных; UN; UNFT; A_UNFT; форматы представления атрибутов объектов (число, время, дата, текст и т. п.).

Справочник атрибутов объектов класса «Форматы»:

A_UNFT.1: тип представления данных; ; N_A_UNFT.1; S ;

A_UNFT.2: вид формата; ; N_A_UNFT.2; S

A_UNFT.3: обязательное (r) или факультативное (f) значение атрибута; ;
N_A_UNFT.3; S ;

A_UNFT.4: уникальное (u) или множественное (m) значение атрибута; ;
N_A_UNFT.4; S

A_UNFT.5: ограничения по структуре; ; N_A_UNFT.5; D

A_UNFT.6: ссылка на описание формата; ; N_A_UNFT; D

Фрагменты словарей значений атрибутов объектов класса «Форматы»:

N_A_UNFT.1.1: текст

N_A_UNFT.1.2: изображение

N_A_UNFT.1.3: видео

N_A_UNFT.1.4: звук

N_A_UNFT.1.5: время в формате чч[.мин[.сек]]

N_A_UNFT.1.6: дата в формате гggg[.мм[дд]]

N_A_UNFT.1.7: формула

N_A_UNFT.1.8: связь

N_A_UNFT.1.9: ссылка на сетевой ресурс

И т. д.

N_A_UNFT.2.1: MP3

N_A_UNFT.2.2: PDF

N_A_UNFT.2.3: Excel, csv

N_A_UNFT.2.4: JPG

N_A_UNFT.2.5: MP4

N_A_UNFT.2.6: простая связь URNc первого типа между объектами, атрибутами или значениями O1 и O2 вида <URNc>:<URNO1><URNO2>, где URNc – URN конкретной связи. Пример: фамилия «Крылов» эквивалентна «Krylov» и «Krylow»; отдел входит в состав института и т. д.

N_A_UNFT.2.7: простая связь второго типа, указывающая на субъект, объект, URN связи и URN значения связи. Формат представления связи имеет вид:

<URNc>:<URN субъекта><URN объекта>=<URN элемента словаря значений соответствующего атрибута связи>. Пример: персона Р1 является сотрудником организации О1 в должности инженера (значение атрибута).

И т. д.

N_A_UNFT.3.1: r

N_A_UNFT.3.2: f

N_A_UNFT.4.1: u

N_A_UNFT.4.2: m

N_A_UNFT.5.1: арабские цифры

N_A_UNFT.5.2: библиографическое описание по ГОСТ
<https://docs.cntd.ru/document/1200034383>

N_A_UNFT.5.3: структура адреса электронной почты – буквенно-цифровая строка, содержащая внутри символ «@», по крайней мере, одну точку не в начале и не в конце строки и не содержащая спецсимволов.

N_A_UNFT.5.4: только буквы

N_A_UNFT.5.5: структура URL – строка символов, начинающаяся с http:// или
<https://>

И т. д.

N_A_UNFT.6.1: <https://ru.wikipedia.org/wiki/JPEG>,
<https://open-file.ru/types/jpeg> [описание формата JPEG]

N_A_UNFT.6.2: <https://ru.wikipedia.org/wiki/MPEG-4>,
<https://open-file.ru/types/mp4> [описание формата mp4]

И т. д.

Фрагмент словаря форматов:

UNFT.1: N_A_UNFT.1.1; ; N_A_UNFT3.1; N_A_UNFT.4.1 [текст, атрибут обязательный, значение уникальное]

UNFT.2: N_A_UNFT.1.7; ; N_A_UNFT3.1; N_A_UNFT.4.1 [дата в формате гггг[.мм[.дд]], атрибут обязательный, значение уникальное]

UNFT.3: N_A_UNFT.1.15; ; N_A_UNFT.3.2; N_A_UNFT.4.2; N_A_UNFT.5.5 [ссылка на внешний ресурс, атрибут необязательный, повторяющийся]

UNFT.4: N_A_UNFT.1.1; ; N_A_UNFT3.1; N_A_UNFT.4.2 [текст, только буквы, атрибут обязательный, значение повторяющееся]

UNFT.5: N_A_UNFT.1.1; ; N_A_UNFT3.2; N_A_UNFT.4.1 [текст, атрибут факультативный, значение уникальное]

UNFT.6: N_A_UNFT.1.1; ; N_A_UNFT3.2; N_A_UNFT.4.2 [текст, атрибут факультативный, значение повторяющееся]

UNFT.7: N_A_UNFT.1.1; ; N_A_UNFT3.1; N_A_UNFT.4.1; N_A_UNFT.5.2 [библиографическое описание элемент обязательный, значение уникальное]

UNFT.8: N_A_UNFT.1.9; ; N_A_UNFT3.2; N_A_UNFT.4.2; N_A_UNFT.5.5 [ссылка на интернет-ресурс, элемент факультативный, значение повторяющееся].

Class.2: Единицы измерения; UN; UNMU; A_UNMU; стандартные единицы измерения различных физических величин.

Справочник атрибутов объектов класса «Единицы измерения»:

A_UNMU.1: наименование единицы измерения; UNFT.1; N_A_UNMU.1; S

A_UNMU.2: предмет измерения; UNFT.4; N_A_UNMU.2; S

A_UNMU.3: обозначения (аббревиатура); UNFT.4; N_A_UNMU.3; S

A_UNMU.4: дополнительная информация; UNFT.18; N_A_UNMU.5; S

Фрагменты словарей значений атрибутов класса «Единицы измерения»:

N_A_UNMU.1.1: секунда

N_A_UNMU.1.2: метр

N_A_UNMU.1.3: грамм

N_A_UNMU.1.4: ньютон

И т. д.

N_A_UNMU.2.1: время

N_A_UNMU.2.2: длина

N_A_UNMU.2.3: масса

N_A_UNMU.2.4: сила

И т. д.

N_A_UNMU.3.1: с.

N_A_UNMU.3.2: сек.

N_A_UNMU.3.3: м.

N_A_UNMU.3.4: г.

N_A_UNMU.3.5: дж.

N_A_UNMU.3.6: н.

И т. д.

N_A_UNMU.4.1: Определение секунды в Википедии

<https://ru.wikipedia.org/wiki/Секунда>

N_A_UNMU.4.2: Определение секунды в Большой российской энциклопедии <https://bigenc.ru/physics/text/3546123>

И т. д.

Элемент словаря объектов класса «Единицы измерения», относящийся к секунде, будет иметь вид:

UNMU.1: N_A_UNMU.1.1; N_A_UNMU.2.1; N_A_UNMU.3.1; N_A_UNMU.3.2; N_A_UNMU.4.1; N_A_UNMU.4.2.

Class.3: Персоны; UN; UNPS; A_UNPS; ; информация о персонах, в той или иной мере связанных с научными исследованиями.

Фрагмент справочника атрибутов объектов класса «Персоны»:

A_UNPS.1: фамилия; UNFT.1; N_A_UNPS.1; D; фамилия выбирается из словаря, при отсутствии она вводится и проверяется на эквивалентность с другими написаниями

A_UNPS.4: дата рождения; UNFT.2; N_A_UNTC.2; D; дата выбирается из словаря временных характеристик, при отсутствии вводится в словарь в соответствии с указанным форматом

A_UNPS.8: квалификация (ученая степень); UNFT.6; N_A_UNPS.8; S; выбирается из словаря;

Фрагмент статического словаря значений атрибута «квалификация (ученая степень)»:

N_A_UNPS.8.1: доктор физ.-мат. наук

N_A_UNPS.8.2: кандидат техн. наук

N_A_UNPS.8.3: PhD в области химии

Class.4: Публикации; UN; UNPB; A_UNPB; книги, сборники, журналы, статьи и т. п.

A_UNPB.1: наименование; UNFT.1; N_A_UNPB.1; D

A_UNPB.2: год издания; UNFT.1; N_UNND; D

A_UNPB.3: вид публикации; UNFT.1; N_A_UNPB.3; S

A_UNPB.4: библиографическое описание; UNFT.7; N_A_UNPB.4; D

Если издание имеет несколько названий, это указывается в виде универсальной связи «эквивалентно» между элементами словаря N_A_UNPB.1. Это позволит поисковому модулю обрабатывать запросы, включающие элементы всех названий данного издания.

Указание на то, что статья опубликована в данном журнале (сборнике), формируется как связь «входит в состав».

Роли персон и организации в данной публикации указываются в виде специфических связей с соответствующими значениями.

Фрагмент статического словаря значений атрибута «вид публикации»:

N_A_UNPB.3.1: монография

N_A_UNPB.3.2: сборник

N_A_UNPB.3.3: статья

N_A_UNPB.3.4: журнал

N_A_UNPB.3.5: выпуск журнала

Class.5: Квалификационные работы UN; UNDS; A_UNDS; диссертации, авторефераты и т. п.

A_UNDS.1: наименование; UNFT.1; N_A_UNDS.1; D

A_UNDS.2: дата выпуска (защиты) работы; UNFT.2; N_A_UNTC.2; D

A_UNDS.3: вид работы; UNFT.1; N_A_UNDS.3; S

A_UNDS.4: URL полного текста; UNFT.9; N_A_UNDS.4; D заполняется в случае отсутствия текста в ЕЦПНЗ, при наличии полного текста в ЕЦПНЗ формируется специфическая связь «квалификационная работа – документ» или «квалификационная работа – изображение»;

A_UNDS.5: Дополнительная информация; UNFT.5; N_A_UNDS.5; D
Фрагмент статического словаря значений атрибута «Вид работы»:

N_A_UNDS.3.1: диссертация докторская

N_A_UNDS.3.2: диссертация кандидатская

N_A_UNDS.3.3: диссертация PhD

Тема диссертации (номер специальности ВАК, индексы УДК и ГРНТИ и др.) указывается как элементы квазиуниверсальной связи с объектами класса «Универсальные классификационные системы». Роли персон и организаций в подготовке и защите диссертации указываются как элементы именованных специфических связей.

Class.6: Музейные предметы; UN; UNMS; A_UNMS; цифровые копии предметов, хранящихся в музеях

A_UNMS.1: наименование предмета; UNFT.1; N_A_UNMS.1; D; ; ;

A_UNMS.2: наименование музея, где хранится в настоящее время оригинал; UNFT.1; N_A_UNOR.1; D; выбирается из словаря организаций, при отсутствии вводится в словарь;

A_UNMS.3: вид источника поступления; UNFT.1; N_A_UNMS.3; S; ;

A_UNMS.4: дата поступления в музей; UNFT.2; N_A_UNTC.2; D; проверяется по словарю временных характеристик. при отсутствии вводится в словарь

A_UNMS.5: дата обнаружения (создания) объекта; UNFT.2; N_A_UNTC.2; D; проверяется по словарю временных характеристик. при отсутствии вводится в словарь

A_UNMS.6: описание музейного предмета; UNFT.5; N_A_UNMS.6; D

Фрагменты статического словаря значений атрибута «Вид источника поступления»

N_A_UNMS.3.1: приобретен музеем

N_A_UNMS.3.2: получен в дар

N_A_UNMS.3.3: изготовлен для музея

Именованные связи музейных предметов с персонами, организациями, событиями и другими классами объектов отражаются в справочниках и словарях соответствующих специфических связей.

ЗАКЛЮЧЕНИЕ

Использование предложенной модели онтологии ЕЦПНЗ позволяет унифицировать алгоритмы создания контента пространства, разработать типовой интерфейс добавления новых элементов вне зависимости от подпространства и конкретного вида данных, упростить и ускорить алгоритмы поиска и навигации по связанным данным. В настоящее время в МСЦ РАН ведутся исследования по развитию и конкретизации предложенной модели в части алгоритмизации формирования вложенных связей, а также моделирования формирования фрагментов ЕЦПНЗ на примере реальных данных, в том числе, составляющих контент электронной библиотеки «Научное наследие России».

Работа выполнена в Межведомственном суперкомпьютерном центре РАН в рамках государственного задания по теме FNEF-2024-0014.

СПИСОК ЛИТЕРАТУРЫ

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.
<https://doi.org/10.31857/S0869-5873897728-735>.
2. Савин Г.И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России. 2020. № 5. С. 3–5.
<https://doi.org/10.51218/0204-3653-2020-5-3-5>.
3. Каленов Н.Е., Сотников А.Н. Архитектура единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 5–8. <https://doi.org/10.51218/0204-3653-2020-5-5-8>.

4. *Абрамов А.Г., Гончар А.А., Евсеев А.В.* Национальная исследовательская компьютерная сеть нового поколения как инфраструктурно-сервисная платформа Единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 43–46.

<https://doi.org/10.51218/0204-3653-2020-5-43-46>.

5. *Соболевская И.Н.* Об особенностях представления мультимедийных объектов в едином цифровом пространстве научных знаний // Информационные ресурсы России. 2020. № 5. С. 31–34.

<https://doi.org/10.51218/0204-3653-2020-5-31-34>.

6. *Sobolevskaya I.* Some Aspects of 3D-objects Presentation in a Common Digital Space of Scientific Knowledge // CEUR Workshop Proceedings (CEUR-WS.org). 2021. Vol. 2990. P. 117–124.

<https://doi.org/10.51218/1613-0073-2990-117-124>.

7. *Каленов Н.Е., Сотников А.Н.* О структуре онтологии Единого цифрового пространства научных знаний // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции. 2022. С. 203–221.

<https://doi.org/10.20948/abrau-2022-23>.

8. *Каленов Н.Е., Сотников А.Н.* Унифицированное представление онтологии единого цифрового пространства научных знаний // Электронные библиотеки. 2023. Т. 26. № 1. С. 80–103.

<https://doi.org/10.26907/1562-5419-2023-26-1-80-103>.

9. *Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N.* Common Digital Space of Scientific Knowledge Ontology Structurization // Lobachevskii Journal of Mathematics. 2023. Vol. 44. No. 7. P. 2733–2743. <https://doi.org/10.1134/S1995080223070235>

10. *Погорелко К.П.* Динамика использования электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сб. науч. тр. 2017. С. 192–200.

11. *Pogorelko K.* A New Version of the Software for the Information System “Scientific Heritage of Russia” // CEUR Workshop Proceedings (CEUR-WS.org). 2021. Vol. 2990. P. 110–116. <https://doi.org/10.51218/1613-0073-2990-110-116>.

12. *Каленов Н.Е., Погорелко К.П., Сотников А.Н.* О развитии электронной библиотеки «Научное наследие России» как составляющей Единого цифрового

пространства научных знаний // Информационные процессы. 2022. Т. 22. № 3. С. 155–166. https://doi.org/10.53921/18195822_2022_22_3_155.

ONTOLOGY OF THE UNIVERSAL SUBSPACE OF COMMON DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE

S. A. Vlasova¹ [0000-0003-1533-5850], **N. E. Kalenov**² [0000-0001-5269-0988],

A. N. Sotnikov³ [0000-0002-0137-1255]

¹⁻³ Joint Supercomputer Center of the Russian Academy of Sciences – JSCC

¹*vlas.svetlana2013@yandex.ru*, ²*nekalenov@mail.ru*, ³*asotnikov@jssc.ru*

Abstract

The work is a development of research conducted by the authors in the field of creating a Common Digital Space for Scientific Knowledge (CDSSK). Previous research has proposed a unified structure for representing the ontology of CDSSK elements (subspaces, classes and attributes of objects, relationships between objects or attributes). In the process of modeling the ontology using the example of the universal and a number of thematic subspaces of the CDSSK, the need for some adjustments to the structure of the ontology regarding CDSSK directories was revealed to ensure the possibility of describing nested data attributes. In addition, the concept of “data attribute value dictionary type” was introduced into the ontology; two types of dictionaries were defined – “static” and “dynamic”. This information makes it possible to simplify formal-logical control algorithms when generating CDSSK content. An indication of the dictionary type has been introduced into the structure of object attribute directories. The presented work describes the modified structure of the ontology using the example of 11 auxiliaries and 10 subject classes of the CDSSK universal subspace (USS). Examples of directories of each class, built in accordance with the ontology structure model, a list of object attributes and examples of static dictionaries are given.

Keywords: digital space of scientific knowledge, ontologies, structuring, linked data.

REFERENCES

1. *Antopol'skiy A.B., Kalenov N.Ye., Serebryakov V.A., Sotnikov A.N.* // O yed-inom tsifrovom prostranstve nauchnykh znaniy // Vestnik Rossiyskoy akademii nauk. 2019. T. 89. № 7. S. 728–735.
2. *Savin G.I.* Edinoe cifrovoe prostranstvo nauchny`x znaniy: celi i zadachi // Informacionny`e resursy` Rossii. 2020. № 5. S. 3–5. <https://doi.org/10.51218/0204-3653-2020-5-3-5>.
3. *Kalenov N.Ye., Sotnikov A.N.* Arkhitektura yedinogo tsifrovogo pros-transtva nauchnykh znaniy // Informatsionnyye resursy Rossii. 2020. № 5. S. 5–8. <https://doi.org/10.51218/0204-3653-2020-5-5-8>.
4. *Abramov A.G., Gonchar A.A., Yevseyev A.V.* Natsional'naya issle-dovatel'skaya komp'yuternaya set' novogo pokoleniya kak infrastruktarno-servisnaya platforma Yedinogo tsifrovogo prostranstva nauchnykh znaniy // Informatsionnyye resursy Rossii. 2020. № 5. S. 43–46.
5. *Sobolevskaya I.N.* Ob osobennostyakh predstavleniya mul'timediynykh ob"yektov v yedinom tsifrovom prostranstve nauchnykh znaniy // Informatsionnyye resursy Rossii. 2020. № 5. S. 31–34. <https://doi.org/10.51218/0204-3653-2020-5-31-34>.
6. *Sobolevskaya I.* Some Aspects of 3D-objects Presentation in a Common Digital Space of Scientific Knowledge // CEUR Workshop Proceedings (CEUR-WS.org). 2021. Vol. 2990. P. 117–124. <https://doi.org/10.51218/1613-0073-2990-117-124>.
7. *Kalenov N.Ye., Sotnikov A.N.* O strukture ontologii Yedinogo tsifrovogo prostranstva nauchnykh znaniy // Nauchnyy servis v seti Internet: trudy XXIV Vse-rossiyskoy nauchnoy konferentsii. 2022. S. 203–221. <https://doi.org/10.20948/abrau-2022-23>.
8. *Kalenov N.Ye., Sotnikov A.N.* Unifitsirovannoye predstavleniye ontologii yedinogo tsifrovogo prostranstva nauchnykh znaniy // Elektronnyye biblioteki. 2023. T. 26, № 1. S. 80–103. <https://doi.org/10.26907/1562-5419-2023-26-1-80-103>.
9. *Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N.* Common Digital Space of Scientific Knowledge Ontology Structurization // Lobachevskii Journal of Mathematics. 2023. Vol. 44. No. 7. P. 2733–2743. <https://doi.org/10.1134/S1995080223070235>.

10. *Pogorelko K.P.* Dinamika ispol'zovaniya elektronnoy biblioteki "Nauchnoye naslediyе Rossiі" // Informatsionnoye obespecheniye nauki: novyye tekhnologii: Sb. nauch. tr. 2017. S. 192–200.

11. *Pogorelko K.* A New Version of the Software for the Information System "Scientific Heritage of Russia" // CEUR Workshop Proceedings (CEUR-WS.org). 2021. Vol. 2990. P. 110–116. <https://doi.org/10.51218/1613-0073-2990-110-116>.

12. *Kalenov N.Ye., Pogorelko K.P., Sotnikov A.N.* O razvitii elektronnoy biblioteki "Nauchnoye naslediyе Rossiі" kak sostavlyayushchey Yedinogo tsifrovogo prostvanstva nauchnykh znaniy // Informatsionnyye protsessy. 2022. T. 22, № 3. S. 155–166. https://doi.org/10.53921/18195822_2022_22_3_155.

СВЕДЕНИЯ ОБ АВТОРАХ



ВЛАСОВА Светлана Александровна – ведущий научный сотрудник Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», кандидат технических наук.

Svetlana Aleksandrovna VLASOVA – Leading Researcher of Joint Super Computer Center of the Russian Academy of Sciences – Branch of Federal State Institution «Scientific Research Institute for System Analysis of the Russian Academy of Sciences», Candidate of Technical Sciences.

email: vlas.svetlana2013@yandex.ru

ORCID: 0000-0003-1533-5850



КАЛЕНОВ Николай Евгеньевич – главный научный сотрудник Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», доктор технических наук, профессор.

Nikolay Evgenievich KALENOV – Chief Researcher of Joint Super Computer Center of the Russian Academy of Sciences – Branch of Federal State Institution «Scientific Research Institute for System Analysis of the Russian Academy of Sciences», Doctor of Technical Sciences, Professor.

email: nekalenov@mail.ru

ORCID: 0000-0001-5269-0988



СОТНИКОВ Александр Николаевич – заместитель директора Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», доктор физико-математических наук, профессор.

Alexander Nikolaevich SOTNIKOV – Deputy Director of the Joint SuperComputer Center of the Russian Academy of Sciences – Branch of the Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, Doctor of Sciences (Math), Professor.

email: asotnikov@jssc.ru

ORCID: 0000-0002-0137-1255

Материал поступил в редакцию 3 декабря 2023 года

УДК 01+002.53

ПЕРЕЧЕНЬ ВАК: ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ В БАЗЕ РНЖ И ELIBRARY.RU

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

Перечень рецензируемых научных журналов ВАК постепенно превращается в достаточно сложную информационную систему, завязанную на нормативные документы ВАК, библиометрические данные eLibrary.ru, решения экспертных советов ВАК и рабочих групп, занимающихся анализом, ранжированием и категорированием журналов Перечня. Созданная ФГБУ РИЭПП база Российские научные журналы (РНЖ) может стать системой, обслуживающей запросы разных категорий пользователей, связанных с темой защиты диссертаций. Пока в РНЖ реализованы интерфейс представителя редакции журнала и интерфейс члена экспертного совета ВАК. В РНЖ желательно включить открытый интерфейс, обращенный соискателю ученой степени, для проверки выполнения требований ВАК к публикациям в журналах из Перечня. При налаженном взаимном обмене данными между РНЖ и eLibrary.ru интерфейс соискателя с обозначенным функционалом может быть организован в среде пользователя eLibrary.ru.

Ключевые слова: научный журнал, информационная система, библиографическая база, Перечень ВАК, база РНЖ, eLibrary.ru, интерфейс соискателя ученой степени

ВВЕДЕНИЕ

Напомним, что представляет собой Перечень рецензируемых изданий (далее – Перечень). В основополагающих документах Высшей аттестационной комиссии (ВАК) при Министерстве науки и высшего образования РФ, в том числе Положении «О порядке присуждения ученых степеней», указано, что для защиты

диссертации на соискание ученой степени соискатель должен иметь статьи, опубликованные в Перечне. Первый вариант Перечня, сформированный в основном силами экспертных советов ВАК, появился в начале 2000-х годов. Долгое время Перечень представлял собой упорядоченный по алфавиту список журналов. Этот список был единственным документом, из которого редакции журналов, а также соискатели ученых степеней могли узнать, обладает ли журнал статусом «входит в Перечень ВАК».

Перечень ВАК с некоторой периодичностью подвергался переработке, в частности, туда добавлялись новые журналы. Сейчас число журналов в Перечне перевалило за 2,5 тысячи.

В 2015 г. ВАК изменил стратегию формирования Перечня. Журнал мог инициативно подать заявку на включение в Перечень, сообщая о себе довольно представительный объем сведений. Для журнала можно было также указать тематику в терминах *научных областей* действующей на тот момент номенклатуры специальностей ВАК. В письме Департамента аттестации научных и научно-педагогических работников Минобрнауки № 13-3337 от 02.07.2015 г. [1] было указано, что с 1 декабря 2015 г. издания, входившие в Перечень 2010 г., но не вошедшие в Перечень 2015 г., не будут признаваться рецензируемыми научными изданиями, в которых должны публиковаться основные научные результаты диссертационных работ.

В 2018 г. ВАК в своем официальном письме рекомендовал журналам из Перечня указать, по каким *специальностям* номенклатуры ВАК журналы публикуют статьи. Если журнал не подавал свои предложения, то привязка к специальностям могла осуществляться экспертными советами ВАК.

В 2021 г. были изданы нормативные документы ВАК о вступлении в действие новой номенклатуры специальностей. В соответствии с рекомендациями Президиума ВАК в июне 2022 г. редакции журналов, входящих в Перечень, подали в ВАК сведения о привязке журналов к специальностям новой номенклатуры.

ОТ СПИСКА К ИНФОРМАЦИОННОЙ СИСТЕМЕ

Еще не так давно актуальный Перечень можно было легко найти на старом сайте ВАК в разделе, содержащем справочные материалы. Сейчас, когда сайт ВАК

разместился на портале Минобрнауки, найти последнюю версию Перечня стало не просто. Если обратиться к поиску в Яндексе, то результаты поискового запроса не обрадуют посетителя: Яндекс выдает адреса десятков сайтов, где размещены разнообразные версии Перечня ВАК. Однако понять, на каком сайте отражена актуальная версия, сейчас довольно сложно.

На сайте Института патологии, фармакологии и терапии [2] размещены версии Перечня, начиная с версии 2001–2005 гг. и включая версию Перечня от 25.04.2023 г. (Рис. 1). Но поскольку этот сайт не является официальным публикатором Перечня, размещенную на сайте информацию можно считать только ориентирующей.

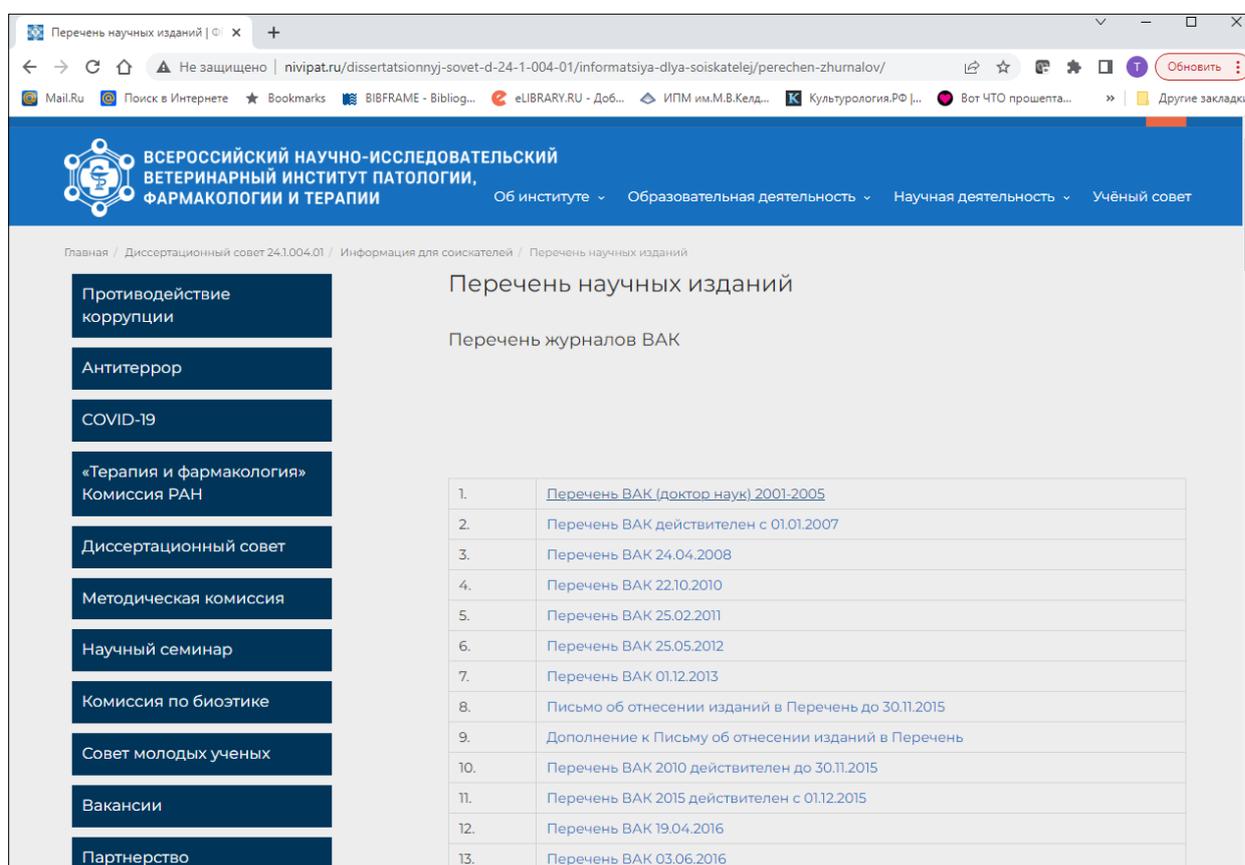


Рис. 1. Ссылки на версии Перечня ВАК на сайте Института патологии, фармакологии и терапии

В феврале-марте 2023 года ВАК запустил кампанию по сбору подробной информации о выпусках журналов из Перечня ВАК за 2022 год, распространив

письмо от 10 февраля 2023 г. № 4/3-разн «О заполнении данных в личных кабинетах журналов Перечня ВАК» [3]. Сбор информации проводился в специально созданной ФГБУ РИЭПП базе «Российские научные журналы» (РНЖ) [4]. Особенность структуры данных базы РНЖ состояла в том, что каждая журнальная статья, опубликованная в журнале в 2022 г., привязывалась к специальностям новой номенклатуры ВАК.

Последние события, связанные с появлением базы РНЖ, дают основания ожидать больших перемен в политике опубликования результатов диссертационных исследований. Вероятно, вскоре могут появиться новые редакции нормативных документов, которые ужесточат требования к опубликованию статей соискателей ученых степеней в профильных журналах из Перечня.

Перечень перестает быть только обычным списком журналов: появилась информационная система РНЖ, содержащая подробные данные о журналах, входящих в Перечень ВАК. Хочется надеяться, что создание и наполнение РНЖ сведениями о выпусках журналов 2022 г. не являются разовой акцией, не имеющей продолжения. По-видимому, ВАК инициировал долгосрочный проект и будет поддерживать эту базу в актуальном состоянии.

Нужно ли хранить, наряду с актуальной на сегодняшний день версией Перечня, информацию о более ранних версиях Перечня? Безусловно, да. В этом случае соискатель ученой степени при наличии соответствующего интерфейса сможет создать в системе РНЖ запрос: входил ли журнал, в котором опубликована его статья, например, в 2016 году, в актуальную на тот момент версию Перечня. База РНЖ, при сохранении информации о более ранних версиях, получит возможность быстро осуществить такую проверку. В число пользователей РНЖ могут быть включены и другие представители научной общественности, тем или иным образом связанные с темой защиты диссертаций: база РНЖ сможет обслуживать запросы широкого круга пользователей на проверку соблюдения требований ВАК к опубликованию результатов диссертационных исследований.

НОВАЯ МЕТОДИКА ОЦЕНКИ ЖУРНАЛОВ ПЕРЕЧНЯ

В письме ВАК от 6 декабря 2022 г. № 02-1198 «О Перечне рецензируемых научных изданий» [5] сообщается, что экспертной группой ВАК утверждена новая методика оценки изданий, входящих в Перечень. Указанная методика включает

в себя две составляющие: количественную, основанную на общепринятых наукометрических показателях, и экспертную. Эксперты оценивают такие показатели журнала, как качество научных статей, авторитетность организации-учредителя, эффективность рецензирования, профессиональный уровень (респектабельность) авторов.

Экспертиза проводится в среде системы РНЖ, где эксперты проходят регистрацию и авторизацию. Размещена инструкция [6] по проведению экспертной оценки на портале РНЖ. Эксперты располагают ориентировочной таблицей, помогающей оценить качественные характеристики издания. В таблице для каждого критерия указан максимально возможный балл. Рассмотрим подробнее критерии, приведенные в инструкции.

Качество научных статей: оценивается логика написания статей, форма изложения, научный уровень публикуемых статей (до 5 баллов);

Уникальность научных статей: оценивается вклад в научную область – постановка новых проблем, добавление новых знаний, выявление ранее неизвестных свойств и т. п. (до 10 баллов);

Уровень авторитетности авторов: оцениваются авторитет, известность в научном сообществе России и/или за рубежом авторов, наличие авторов, обладающих учеными степенями (до 10 баллов);

Качество организации рецензирования: оцениваются политика рецензирования, число рецензентов, срок рецензирования, научный уровень рецензентов (до 5 баллов);

Организация-учредитель: оцениваются наличие научной школы в организации-учредителе, организационно-правовая форма организации, связи с другими научными и образовательными организациями, состав редакционной коллегии (до 5 баллов);

Также оценивается соответствие содержания журнала заявленным специальностям. Это качество оценивается на основании введенной информации о специальностях статей журнала. Баллы получаются автоматически. Если от 1% до 15% статей опубликовано по специальностям журнала, то журнал получает 1 балл; если 16%–30% статей по специальностям журнала – 2 балла, 31%–45% – 3 балла, 46%–59% – 4 балла, от 60% – 5 баллов.

Помимо выставления баллов эксперт имеет возможность дать развернутое пояснение своей позиции, которое затем отразится в экспертном заключении. Эксперт дает также рекомендации по присвоению журналу категории (К1, К2, К3).

В упомянутом письме ВАК от 6 декабря 2022 г. № 02-1198 все научные журналы, входящие в Перечень, были ранжированы и распределены по категориям К1, К2, К3 в соотношении: К1 – 25%, К2 – 50%, К3 – 25% журналов. В приложении к письму ВАК приведен список журналов Перечня. В списке содержатся 2587 наименований журналов с указанием категорий. В письме ВАК также отмечено, что готовятся соответствующие изменения в нормативные акты в сфере государственной научной аттестации. После внесения этих изменений будет определен порядок применения категорирования журналов, входящих в Перечень, при оценке диссертационных работ на предмет удовлетворения требований ВАК к опубликованию результатов диссертационных работ в рецензируемых журналах.

Кроме того, в письме отмечается, что журналы, входящие в международные базы данных Web of Science, Scopus, PubMed, MathSciNet, zbMATH, Chemical Abstracts, Springer или GeoRef, а также журналы RSCI, приравниваются к изданиям категории К1. При этом требование об обязательном наличии у соискателя публикаций в журналах, индексируемых в указанных базах, в настоящее время не применяется.

На сайте ВАК на портале Минобрнауки в разделе «Справочные материалы» [7] есть ссылка на Перечень рецензируемых изданий. Посетитель сайта мог бы ожидать, что здесь ему будет предоставлена последняя версия Перечня. Однако в документе с сайта ВАК нет каких-либо упоминаний о версии Перечня. В Перечне с сайта ВАК содержатся 2269 научных журналов, в то время как в упомянутом выше письме ВАК от 6 декабря 2022 г. № 02-1198 содержатся 2587 журналов. В версии Перечня с сайта ВАК у журналов отсутствуют специальности ВАК и категории К1, К2, К3. Следовательно, сайт ВАК на портале Минобрнауки, вопреки ожиданиям, не является достоверным поставщиком информации о текущей действующей версии Перечня.

В одной из последних версий Перечня от 25.04.2023 г., найденной в интернете на сайтах нескольких вузов, содержатся уже 2919 наименований журналов, что заметно больше, чем в версии Перечня с сайта ВАК. В версии Перечня от 25.04.2023 г. нет разбиений на категории К1, К2, К3. Но есть привязка каждого

журнала к специальностям номенклатуры ВАК. Для каждой специальности указана дата начала действия привязки специальности к журналу. Для специальности из старой номенклатуры указан срок окончания действия привязки специальности к журналу.

По каким причинам в версию Перечня от 25.04.2023 г. не вошли показатели категорий, сказать трудно. Продолжает ли действовать новая методика оценки журналов, упомянутая в письме ВАК от 6 декабря 2022 г.? За период, прошедший с выхода указанного письма, число журналов в Перечне увеличилось с 2587 до 2919. В какую категорию были добавлены более 300 новых журналов? Не изменились ли категории журналов, входящих в Перечень, представленный в письме от 6 декабря 2022 г., в связи с вероятным изменением числа журналов в каждом квартале?

Уже сейчас очевидно, что обилие версий Перечня и динамика включения в Перечень новых журналов требуют создания информационной системы, которая давала бы исчерпывающие ответы на вопросы редакций журналов, диссертационных советов и соискателей. Может ли РНЖ стать такой системой?

БАЗА ЖУРНАЛОВ РНЖ

Рассмотрим подробнее, как осуществлялся сбор информации о выпусках журналов в РНЖ на сайте ФГБУ РИЭПП [8] (Рис. 2). Сотрудник, уполномоченный редакцией журнала, получал доступ к личному кабинету и по специальному запросу модератору выбирал нужный журнал из Перечня.

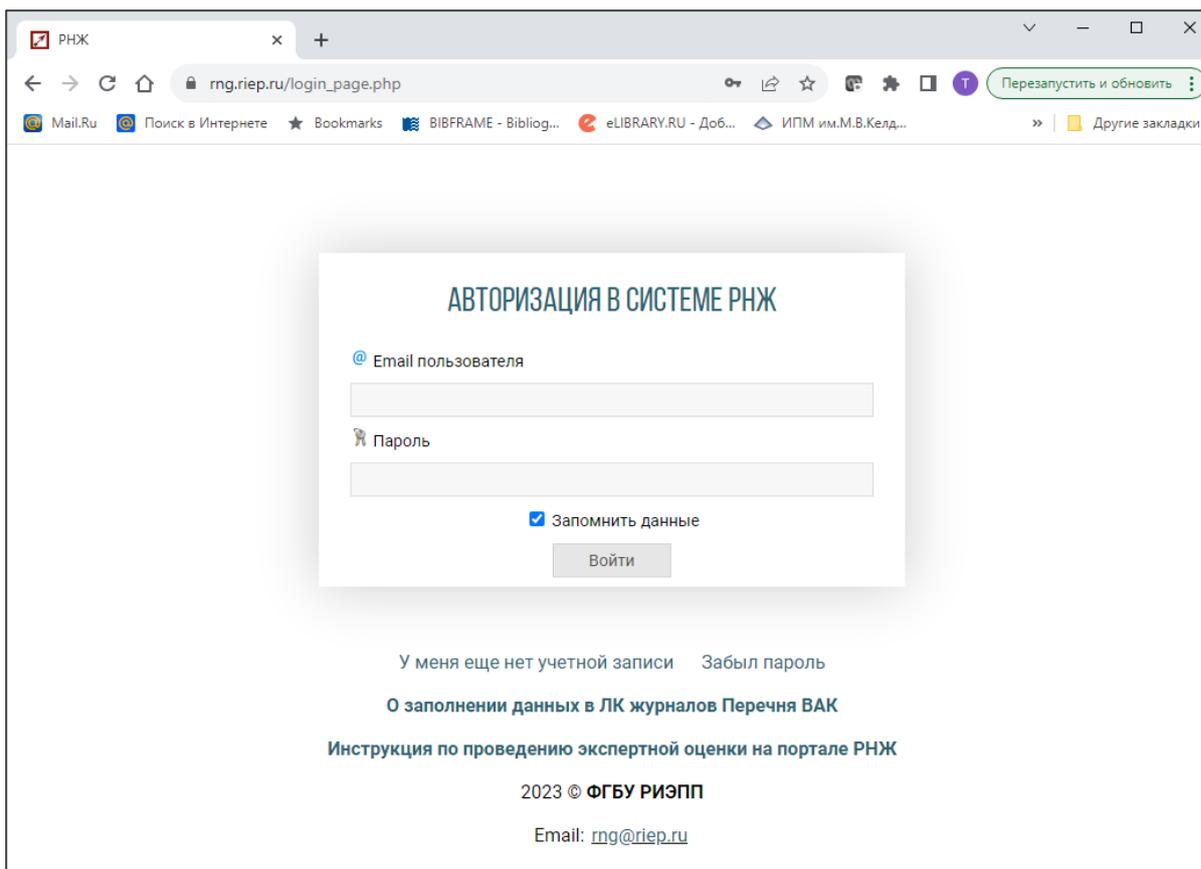


Рис. 2. Вход в личный кабинет представителя издательства на сайте РНЖ ФГБУ РИЭПП

После выбора журнала представитель редакции видит на информационной странице своего журнала в системе РНЖ ряд полей, заполненных системой, с некоторыми формальными параметрами журнала.

Наряду с формальными данными, на странице появлялась следующая информация (Рис. 3):

- категория журнала (К1, К2 или К3) в рейтинге Перечня (поле «категория журнала в рейтинге ВАК»),
- ссылки на версии Перечня, в которых присутствует журнал (поле «Индексация вхождения в списки»).

Представитель редакции должен был ввести недостающую информацию в соответствующие пустые поля. В частности, нужно было привязать журнал к специальностям ВАК. Отметим, что привязка журнала к специальностям контролировалась модератором. Список специальностей должен был, по-видимому, совпадать с официальными документами, отправленными редакцией журнала в ВАК

при регистрации (перерегистрации) специальностей ВАК. На Рис. 3 представлена страница журнала в системе РНЖ с перечисленными выше полями.

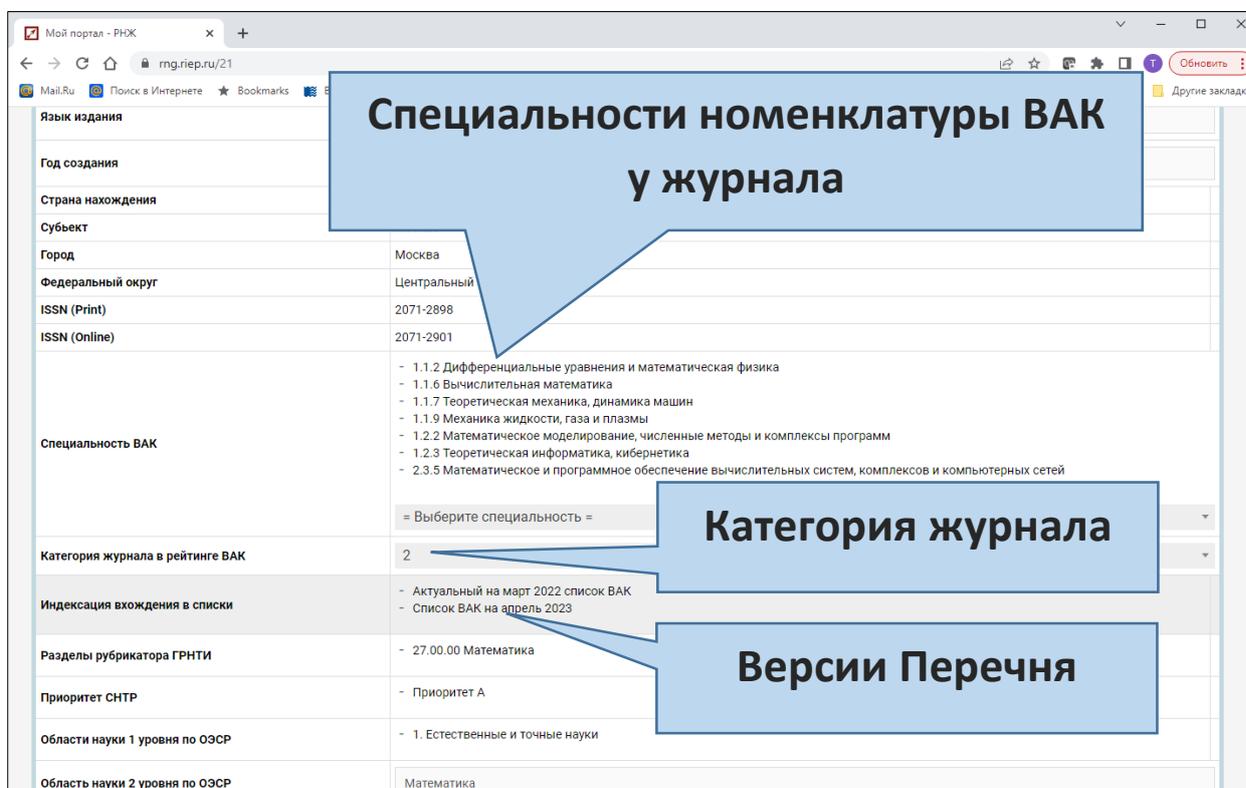


Рис. 3. Страница журнала ВАК, содержащая сведения о специальностях номенклатуры ВАК, категории журнала в рейтинге ВАК, версии Перечня

На Рис. 3 показана страница журнала «Препринты ИПМ им. М.В. Келдыша». К этому журналу привязаны следующие специальности:

- 1.1.2. Дифференциальные уравнения и математическая физика;
- 1.1.6. Вычислительная математика;
- 1.1.7 Теоретическая механика, динамика машин;
- 1.1.9. Механика жидкости, газа и плазмы;
- 1.2.2. Математическое моделирование, численные методы и комплексы программ;
- 1.2.3. Теоретическая информатика, кибернетика;
- 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Журналу «Препринты ИПМ им. М.В. Келдыша» присвоена категория К2. Указано, что журнал входит в следующие версии Перечня (копируем информацию из полей):

- Актуальный на март 2022 список ВАК;
- Список ВАК на апрель 2023.

На странице журнала появились следующие статистические и библиометрические сведения, относящиеся к журналу:

- общее число выпусков,
- число выпусков в год,
- суммарное число цитирований журнала в РИНЦ,
- показатель журнала в рейтинге Science Index,
- двухлетний импакт-фактор,
- двухлетний импакт-фактор по ядру РИНЦ

и т. д.

Наличие этой информации позволяет утверждать, что в системе РНЖ налажен механизм получения данных о журнале из базы eLibrary.ru [9].

Помимо общей информации о журнале в базу РНЖ нужно было загрузить метаданные каждой статьи (авторы, название статьи, аннотация, ключевые слова, библиографический список). На Рис. 4. показана страница с метаданными статьи журнала. Помимо привычных элементов метаданных на странице указана привязка статьи к специальности номенклатуры ВАК. Система РНЖ позволяет привязать к статье до трех специальностей номенклатуры ВАК.

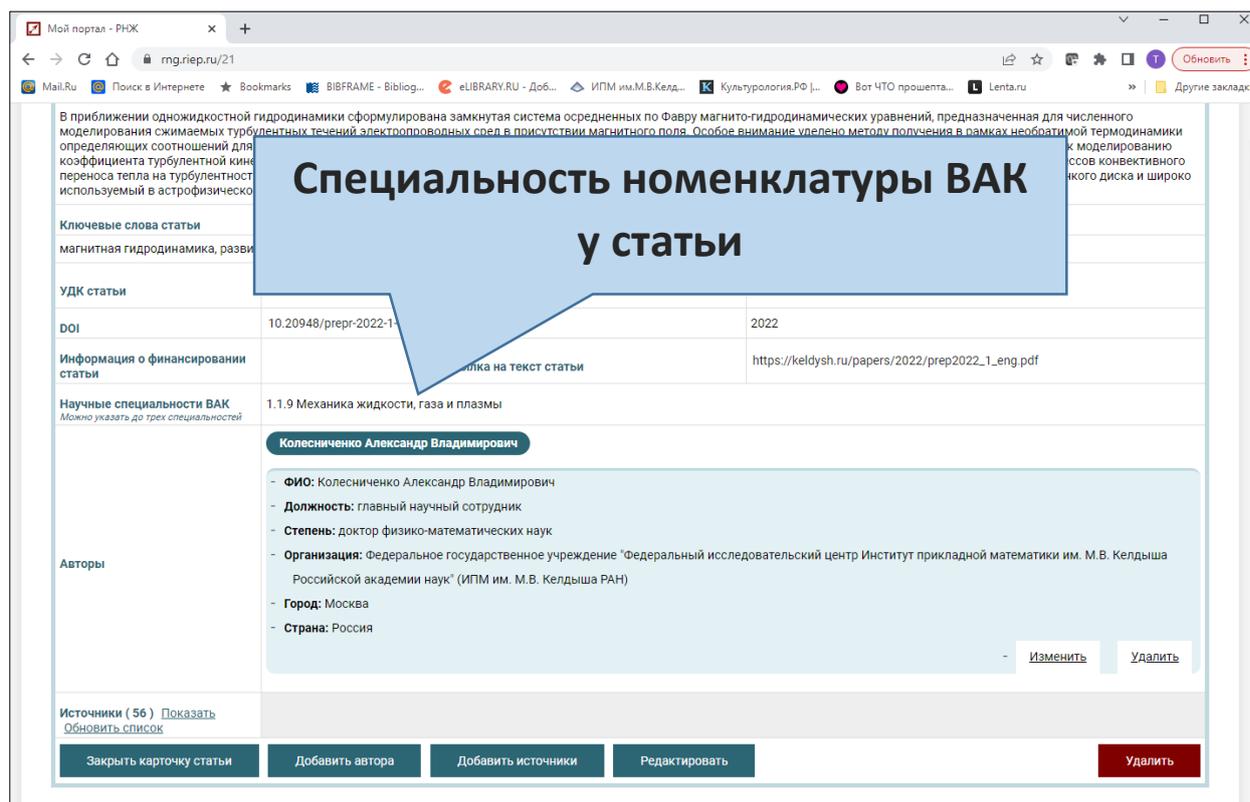


Рис. 4. Метаданные статьи в базе РНЖ

Особенность структуры собираемой информации в системе РНЖ состоит также в том, что собирается довольно подробная информация об авторе статьи (Рис. 4). В частности, указываются следующие параметры:

- должность,
- ученая степень,
- место работы автора,
- город и страна.

База РНЖ предполагает импорт XML-файла с данными о статьях (эта функция в настоящее время проходит тестирование).

ЕДИНОЕ ОКНО

Можно только приветствовать появление информационной системы, связанной с Перечнем. Однако вопросы вызывает очевидное дублирование ввода метаданных выпусков журналов в двух базах: eLibrary.ru и РНЖ.

Журналы из Перечня уже много лет передают метаданные опубликованных статей в базу eLibrary.ru. В eLibrary.ru налажен механизм формирования структуры метаданных, например, с помощью программной среды Articulus. Часть редакций журналов практикует другой канал ввода метаданных – механизм API. Зачем же повторно передавать метаданные в РНЖ? Почему бы не вводить информацию о привязке специальностей номенклатуры ВАК непосредственно в eLibrary.ru через уже действующие механизмы ввода метаданных, расширив программу Articulus и структуру XML в API?

Вероятно, такое решение не состоялось в силу организационных неурядиц. ВАК имеет достаточное влияние на редакции журналов, входящих в Перечень, но этого влияния не хватает для перестройки работы фактически независимой базы eLibrary.ru. Можно ожидать, что со временем две библиографические базы все же наладят конструктивное взаимодействие. Пока редакции журналов вынуждены мириться с временными решениями.

Вводимая в РНЖ информация за 2022 г. в некоторых срезах явно избыточна. Например, в РНЖ предлагают ввести информацию о должности, ученой степени и ученом знании автора. Эта и другая персональная информация уже содержится в eLibrary.ru в профиле автора. Решение eLibrary.ru более технологично, чем ввод в РНЖ подробной информации об авторах каждой статьи. Повторный ввод или дублирование информации, как известно, являются источником ошибок. Данные об авторах в eLibrary.ru контролируются авторами. Изменять авторские данные, следуя логике eLibrary.ru, может только сам автор. В собственном профиле автор с большой вероятностью разместит правильные и актуальные данные.

В настоящее время редакции журналов находятся в сложном положении, поскольку нет каких-либо сведений о том, как долго будут действовать те или иные временные решения в РНЖ. Не исключено, что кампания сбора сведений в РНЖ является разовой акцией. Возможен и другой вариант развития событий: период временных решений будет длиться несколько лет, и в этом случае ежегодная работа редакции по вводу метаданных выпусков журнала в библиографические базы фактически удваивается.

Обратим внимание, что в базу РНЖ уже включены библиометрические данные, передаваемы из eLibrary.ru, т. е. взаимодействие двух баз уже началось. Вопрос состоит в том, будут ли и другие данные (наряду с библиометрическими) из

eLibrary.ru передаваться в РНЖ в переходный период? Если такой обмен данными будет возможен, то это значительно упростит работу представителя журнала: он будет вводить в РНЖ только те данные, которые отсутствуют в eLibrary.ru, например, сведения о привязке статей к специальностям номенклатуры ВАК. Уместной была бы обратная операция, когда данные из РНЖ дополнили бы соответствующие структуры в eLibrary.ru.

Одним из решений, удобных для редакций журналов, могла бы стать разработка единого формата XML-файла с метаданными для РНЖ и eLibrary.ru. При отсылке обобщенного XML-файла в РНЖ или eLibrary.ru каждая из этих двух баз могла бы забирать нужные ей данные. Все вопросы гармонизации двух баз должны быть скрыты от представителей редакций журналов и проводиться в режиме, удобном для разработчиков этих библиографических баз.

Есть еще один поворот темы обновления информации. Многие журналы размещают на своих сайтах некоторые библиометрические данные из eLibrary.ru, например, позицию в тематическом рейтинге и значение параметра рейтинга. Нужен механизм автоматического отображения на сайтах журналов информации о библиометрических показателях. В случае обновления рейтингов eLibrary.ru информация о библиометрических данных обновлялась бы и на сайтах журналов через механизмы кросс-доменного взаимодействия.

ПОЛЬЗОВАТЕЛИ БАЗЫ РНЖ

Можно предположить, что база РНЖ постепенно превратится в специализированную базу, обслуживающую формирование и функционирование Перечня ВАК. Система РНЖ могла бы располагать интерфейсами:

- для ввода/коррекции, мониторинга журналов из Перечня рабочей группой, формирующей актуальный Перечень,
- обслуживания работы экспертных советов ВАК,
- для ввода/коррекции данных о журнале, (интерфейс, ориентированный на уполномоченного представителя журнала).

Как будет организован интерфейс с рабочей группой ВАК, сейчас предсказать довольно сложно. Самый простой вариант взаимоотношений – заказ (в письменной форме на основе нормативных документов) на ввод и коррекцию Перечня персоналу, сопровождающему базу РНЖ.

Интерфейс для экспертов, проводящих экспертизу журналов Перечня, уже реализован на сайте РНЖ (подробности можно найти в инструкции для экспертов [6]).

Интерфейс базы РНЖ, ориентированный на представителя журнала, был представлен выше в разделе «База журналов РНЖ».

К перечисленным интерфейсам можно было бы добавить еще один – открытый интерфейс, которым мог бы воспользоваться любой пользователь интернета. Так, например, соискатель ученой степени мог бы проверить, входит ли журнал, опубликовавший его статью, в соответствующую версию Перечня. Этот вопрос не настолько простой, как кажется на первый взгляд. Журнал может не входить в текущую (последнюю) версию ВАК, но мог входить в Перечень в год опубликования статьи соискателя. Или же журнал мог быть отозван из Перечня за обнаруженные нарушения с потерей высокого статуса «журнал из Перечня ВАК». Важно знать дату «вступления в силу» новой версии Перечня, определяемой нормативными документами ВАК. Только в этом случае можно получить исчерпывающий ответ о вхождении конкретного журнала в конкретный период времени в Перечень.

Возможны и другие события, требующие детального анализа состояния Перечня ВАК. Так, например, в Положение о порядке защиты диссертаций может быть введено требование об обязательной привязке журнала к специальностям из номенклатуры ВАК и учете при защите диссертации только тех статей, которые опубликованы в журнале по соответствующей специальности. Как уже отмечалось, обновленные версии нормативных документов ВАК готовятся. Будут ли учитываться статьи без привязки к специальностям, опубликованные до вступления в силу будущей редакции Положения? Будут ли учитываться статьи, привязанные к близким специальностям номенклатуры ВАК? Для получения ответа на поставленные вопросы нужно дожидаться появления обещанных изменений в нормативных документах ВАК.

Весьма динамичной может оказаться присвоенная журналу категория К1, К2 или К3. Утверждается, что категория зависит от библиометрических показателей журнала, полученных из eLibrary.ru. Известно, что эти показатели регулярно пересчитываются и могут сильно изменяться в случае, например, временной не-

допоставки в eLibrary.ru сведений от редакции журнала. Категория журнала может измениться и по разным неожиданным причинам. Например, в недавнем письме из ВАК от 2 июня 2023 г. № 10/1-разн(об) журналы получили предупреждение, что если они не прошли регистрацию и верификацию на портале РНЖ, то при ранжировании такие журналы будут отнесены к группе журналов категории К3.

Как часто будут пересматриваться категории К1, К2, К3? Будут ли фиксироваться нормативными документами категории журналов на определенный период времени? Будем ждать соответствующих разъяснений ВАК.

ВОЗМОЖНОСТИ ELIBRARY.RU ДЛЯ ВИЗУАЛИЗАЦИИ ПЕРЕЧНЯ

При наличии механизмов взаимодействия и обмена данными между РНЖ и eLibrary.ru интерфейс соискателя (точнее – любого пользователя, который интересуется правилами публикации результатов диссертационного исследования в журналах из Перечня) может быть организован в среде eLibrary.ru.

Например, в eLibrary.ru уже сейчас через стандартный поисковый запрос можно найти журналы, входящие в список ВАК и относящиеся к определенной тематике, задаваемой рубриками ГРНТИ. На Рис. 5 представлена поисковая форма, где можно задать соответствующие параметры (на Рис. 5 задан поиск журналов, входящих в Перечень ВАК и относящихся к теме ГРНТИ «Космические исследования»).

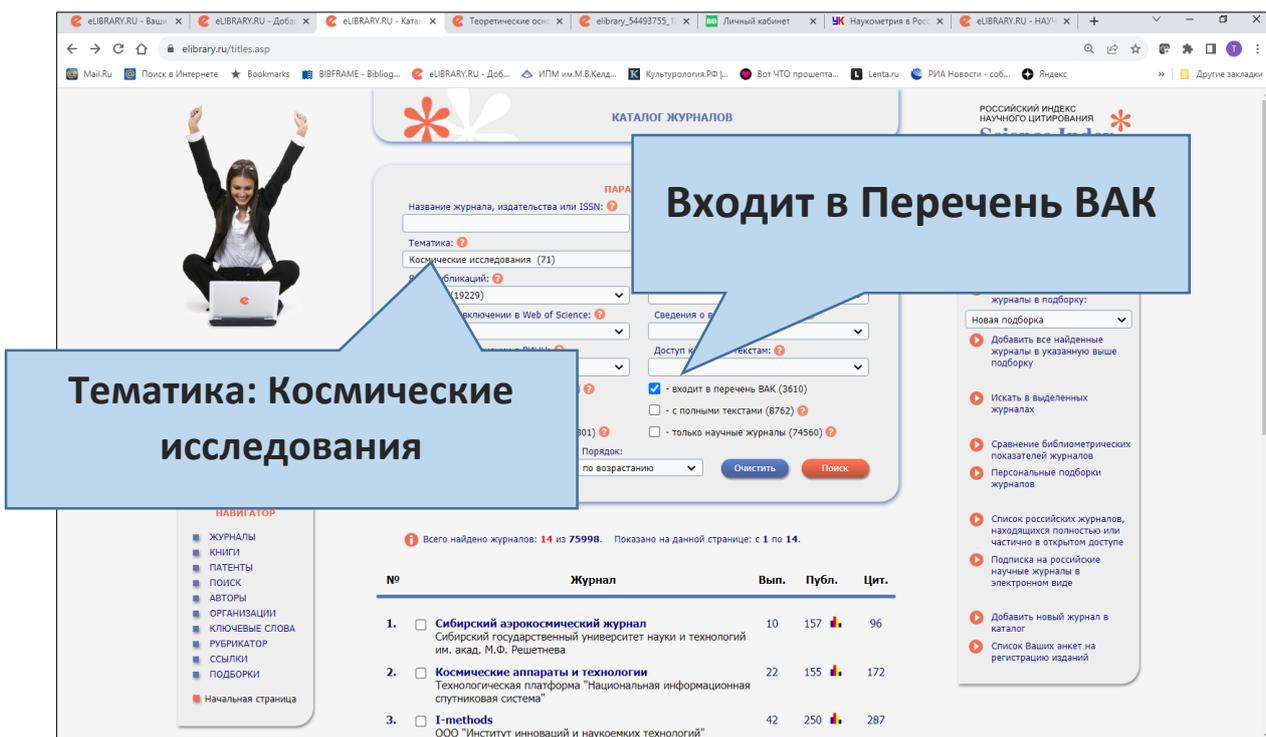


Рис. 5. Форма для поиска журналов по заданным параметрам в eLibrary.ru

Если бы eLibrary.ru получила информацию из РНЖ о привязке журнала к специальностям номенклатуры ВАК, в поисковую форму можно было бы добавить поля для задания специальностей номенклатуры, а также категории журнала в Перечне (К1, К2, К3).

Получив результаты поиска, пользователь может перейти на страницу с анкетой выбранного журнала. Сейчас в анкете журнала указана краткая информации о вхождении журнала в Перечень. На наш взгляд, в анкете журнала в eLibrary.ru должна размещаться более полная информация о журнале, которая вводится в РНЖ.

В настоящее время в анкете указываются рубрики старой номенклатуры ВАК верхнего уровня и краткая информация о вхождении журнала в Перечень ВАК – параметр «Перечень ВАК: да» (Рис. 6). В случае налаженного обмена данными между РНЖ и eLibrary.ru в анкете журнала из Перечня могла бы появиться актуальная информация о статусе журнала – категория, показатель рейтинга, привязанные специальности и т. д.

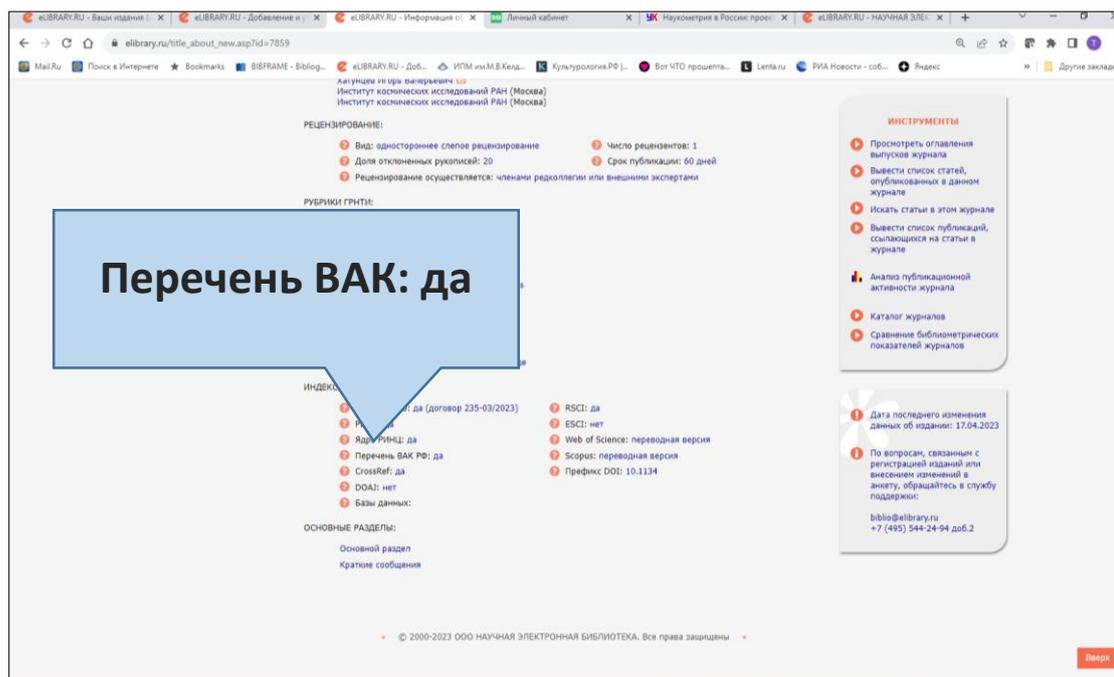


Рис. 6. Анкета журнала в eLibrary.ru

Из анкеты журнала можно перейти к просмотру библиометрических показателей журнала по кнопке «Анализ публикационной активности журнала» и далее перейти на форму просмотра рейтингов eLibrary.ru. На Рис. 7 представлена форма для визуализации рейтингов журналов.

В форме для построения рейтингов должна быть возможность выбора журналов по специальности из номенклатуры ВАК (параметр «Специальность номенклатуры ВАК») – сейчас в качестве тематики можно указать рубрику ГРНТИ.

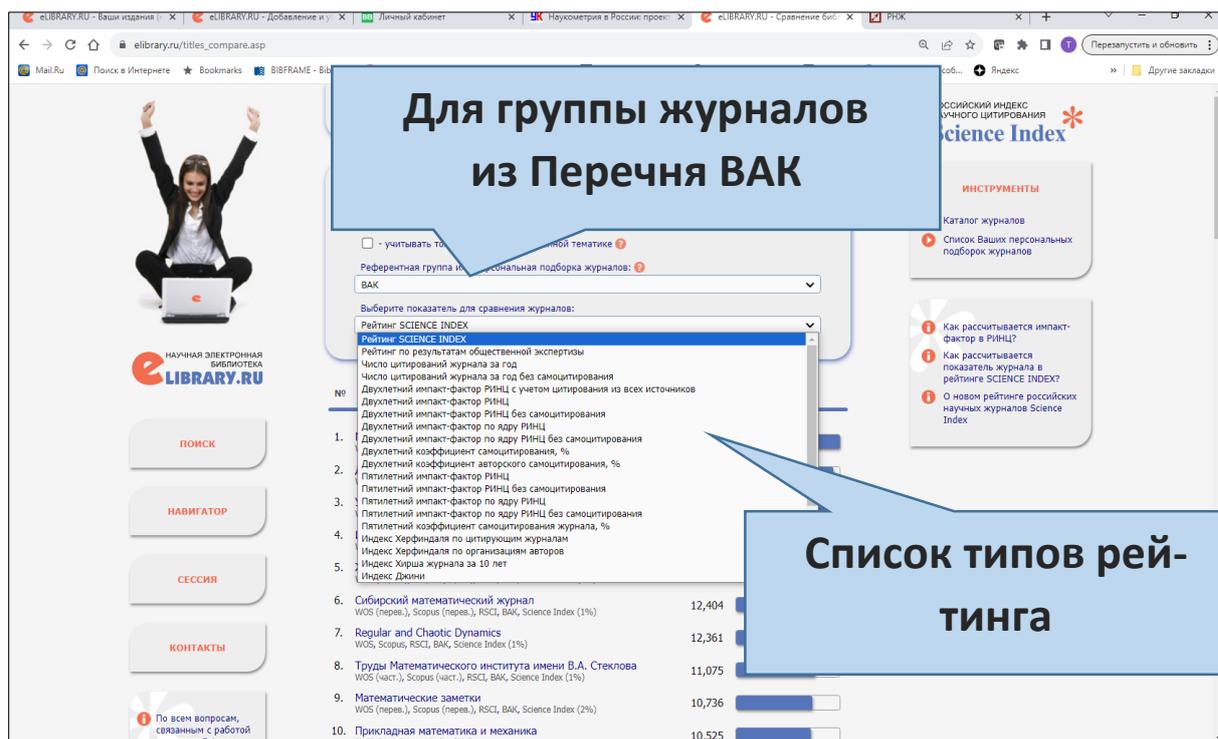


Рис. 7. Форма для визуализации рейтингов журналов в eLibrary.ru

В eLibrary.ru можно построить несколько типов рейтинга. При построении рейтинга можно задать тип, указав его в поле формы. Список типов рейтингов должен быть расширен: в него желательно включить тип «Рейтинг ВАК».

ЗАКЛЮЧЕНИЕ

Перечень журналов ВАК становится достаточно сложной и динамично развивающейся информационной системой, завязанной на нормативные документы ВАК, библиометрические данные eLibrary.ru, решения экспертных советов и рабочих групп ВАК. Перечень уже не может существовать в виде статического документа без идентифицированной версии. Созданная база РНЖ могла бы стать той информационной системой, в которой осуществлялись бы формирование и коррекция состава Перечня, мониторинг и ранжирование входящих в Перечень журналов.

Сейчас в РНЖ реализован интерфейс для сотрудника журнала, входящего в Перечень, и интерфейс для члена экспертного совета ВАК. Необходим также общедоступный интерфейс, позволяющий ответить на вопрос: входил ли журнал, в котором опубликована статья диссертанта, в актуальную версию Перечня на момент публикации статьи с учетом ее специальности.

При налаженном обмене информацией между РНЖ и eLibrary.ru доступ к журналам Перечня для соискателей ученых степеней и широких кругов научной общественности можно реализовать через услуги библиографической базы eLibrary.ru. Перейдя на анкету журнала, пользователь должен увидеть полную информацию о том, когда журнал входил в Перечень и по каким специальностям номенклатуры ВАК.

Перечень может быть показан в eLibrary.ru через форму поиска группы журналов с включенной кнопкой «Входит в Перечень ВАК». В поисковую форму целесообразно добавить поля, в которых можно задавать специальности номенклатуры ВАК, а также категорию журнала К1, К2, К3.

СПИСОК ЛИТЕРАТУРЫ

1. Письмо департамента аттестации научных и научно-педагогических работников Минобрнауки №13-3337 от 02.07.2015 г.

https://spmi.ru/sites/default/files/imci_images/sciens/dissovet/reczurn/VAK/2001-2015/pismo_ob_otnesenii_izdaniy_v_perechen_do_30.11.2015.pdf

2. Версии Перечня ВАК сайте Института патологии, фармакологии и терапии. <http://www.nivipat.ru/dissertatsionnyj-sovet-d-24-1-004-01/informatsiya-dlya-soiskatelej/perechen-zhurnalov/>

3. Письмо от 10 февраля 2023 г. № 4/3-разн «О заполнении данных в личных кабинетах журналов Перечня ВАК». <https://rng.riep.ru/help/recomend.pdf>

4. База «Российские научные журналы» (РНЖ) на сайте ФГБУ РИЭПП. https://rng.riep.ru/login_page.php

5. Письмо ВАК РФ от 6 декабря 2022 г. № 02-1198 «О Перечне рецензируемых научных изданий».

<https://www.garant.ru/products/ipo/prime/doc/405821249/?ysclid=lilhqkkqwg60916187>

6. Инструкция по проведению экспертной оценки. https://rng.riep.ru/help/instruction_expert2.pdf

7. Раздел «Справочные материалы» на сайте ВАК. https://vak.minobrnauki.gov.ru/searching#tab=_tab:materials~ (обращение к сайту 17 октября 2023 г.)

8. Приложение к письму от 10 февраля 2023 г. № 4/3-разн «О заполнении данных в личных кабинетах журналов Перечня ВАК». Инструкция по регистрации на портале РНЖ. <https://rng.riep.ru/help/recomend.pdf>

9. Научная электронная библиотека eLibrary.ru.
<https://www.elibrary.ru/defaultx.asp>

LIST OF HAC: USER INTERFACE IN THE RSJ DATABASE AND ELIBRARY.RU

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

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

polilova@keldysh.ru

Abstract

The List of peer-reviewed scientific journals of the Higher Attestation Commission is gradually turning into a complex information system based on the normative documents of the Higher Attestation Commission, bibliometric data eLibrary.ru, decisions of the expert councils of the Higher Attestation Commission and working groups engaged in the analysis, ranking and categorization of the journals of the List. The Russian Scientific Journals (RSJ) database created by RIEP can become a system that serves the requests of different categories of users related to the topic of dissertation defense. So far, the RSJ has implemented the interface of a representative of the editorial board of the journal and the interface of a member of the expert council of the Higher Attestation Commission. It is desirable to include in the RSJ an open interface addressed to the degree applicant to verify compliance with the requirements of the Higher Attestation Commission for publications in journals from the List. With the established mutual exchange of data between RSJ and eLibrary.ru, the applicant's interface with the designated functionality can be organized in the eLibrary.ru user environment.

Keywords: *scientific journal, information system, bibliographic database, HAC List, RSJ database, eLibrary.ru, the interface of the scientific degree applicant*

REFERENCES

1. Pismo departamenta attestatsii nauchnykh i nauchno-pedagogicheskikh rabotnikov Minobrnauki №13-3337 ot 02.07.2015 g.
https://spmi.ru/sites/default/files/imci_images/sciens/dissovet/reczurn/VAK/2001-2015/pismo_ob_otnesenii_izdanij_v_perechen_do_30.11.2015.pdf
2. Versii Perechnia VAK saite Instituta patologii, farmakologii i terapii.
<http://www.nivipat.ru/dissertatsionnyj-sovet-d-24-1-004-01/informatsiya-dlya-soiskatelej/perechen-zhurnalov/>
3. Pismo ot 10 fevralia 2023 g. № 4/3-razn «O zapolnenii dannykh v lichnykh kabinetakh zhurnalov Perechnia VAK». <https://rng.riep.ru/help/recomend.pdf>
4. Baza «Rossiiskie nauchnye zhurnaly» (RNZh) na saite FGBU RIEPP.
https://rng.riep.ru/login_page.php
5. Pismo VAK RF ot 6 dekabria 2022 g. № 02-1198 «O Perechne retsenziruemykh nauchnykh izdaniy». <https://www.garant.ru/products/ipo/prime/doc/405821249/?ysclid=lilhqkkqwg60916187>
6. Instruktsiia po provedeniiu ekspertnoi otsenki.
https://rng.riep.ru/help/instruction_expert2.pdf
7. Razdel «Spravochnye materialy» na saite VAK.
https://vak.minobrnauki.gov.ru/searching#tab=_tab:materials~ (obrashchenie k сайту 17 oktiabria 2023 g.)
8. Prilozhenie k pismu ot 10 fevralia 2023 g. № 4/3-razn «O zapolnenii dannykh v lichnykh kabinetakh zhurnalov Perechnia VAK». Instruktsiia po registratsii na portale RNZh. <https://rng.riep.ru/help/recomend.pdf>
9. Nauchnaia elektronnaia biblioteka eLibrary.ru.
<https://www.elibrary.ru/defaultx.asp>

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 13 декабря 2023 года

УДК 002.513.5: 004.738

СУБЪЕКТИВНЫЕ ЗАМЕТКИ О ПОИСКОВЫХ СИСТЕМАХ

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

Поводом для данной работы послужили случившиеся один за другим 25-летние юбилеи главных для нашей страны поисковых машин Яндекс и Google. Эта статья – попытка описать некоторые события из истории развития средств навигации в интернете с точки зрения их свидетеля (и частично – участника).

Ключевые слова: Яндекс, Google, поисковые системы, история поиска

ВВЕДЕНИЕ

Поиском информации занимается каждый ежедневно и неоднократно, порой даже не осознавая этого. Наряду с различными формами общения это один из самых популярных видов деятельности в интернете. 10 лет назад на вопрос ФОМ «Что вам приходилось делать в интернете за последний месяц?», 73% интернет-аудитории ответили: «Искать информацию». При этом 64% читали новости, 63% общались в социальных сетях, 45% вели электронную переписку¹. Поиск нужен всегда и везде – в профессиональной деятельности, в обучении, на досуге.

Однако так было не всегда. Люди старшего поколения помнят: ещё 30 лет назад интернет вне научных лабораторий и оборонных спецпроектов был редкой экзотикой. Информационных ресурсов в сети было немного, и не было нужды в сложных системах навигации. Роль справочников играли рукописные списки, позже текстовые файлы. Но в 1993 г. появился браузер Mosaic, о котором изобретатель сети Ethernet Роберт Меткалф высказался так: «Marc Andreessen and Eric Bina developed NCSA Mosaic at the University of Illinois. Several million people then

¹ <http://fom.ru/smi-i-internet/11088>

suddenly noticed that the Web might be better than sex»². Начался взрывной рост числа сайтов, пользователей, объемов информации. Возникла необходимость упорядочивать и искать эту информацию. Стали появляться электронные каталоги веб-адресов – универсальные и специализированные. Но поисковые возможности каталогов ограничивались поиском по самим каталогам, а не по веб-страницам. Вскоре инструменты поиска прошли путь от списков и структурированных каталогов (directories) до мощных поисковых систем (search engines), которые образовали основу крупнейших порталов, стали ведущими трафикогенераторами мирового и российского интернета.

ИЗ РАННЕЙ ИСТОРИИ

В середине 1990-х годов начала складываться современная картина поисковых средств. В 1994 г. стэнфордские аспиранты Д. Фило и Дж. Янг, недовольные имевшимися тогда каталогами, сделали для себя инструмент навигации, из которого выросла база данных Yahoo! (Yet Another Hierarchical Officious Oracle).

Для популярных в то время сетевых протоколов FTP и Gopher появились поисковые программы Archie (1990, A. Emtage) и Veronica (1991, M. McCahill). Archie строил базу для поиска по именам файлов из ftp-архивов, однако не индексировал содержание этих файлов. Veronica (Very Easy Rodent-Oriented Net-wide Index to Computer Archives) выполняла поиск по ключевым словам заголовков меню Gopher и именам файлов [1].

Первой полнотекстовой поисковой системой, индексирующей ресурсы при помощи робота, стала WebCrawler (1993, B. Pinkerton). На момент запуска в ее базе было 4000 веб-сайтов. Она позволяла искать любые слова, расположенные на любой веб-странице. Вскоре появились поисковые машины Lycos (1994), Excite (1995), AltaVista, обрабатывавшая кириллические запросы (1995), Inktomi (1996), HotBot (1996), Northern Light с возможностью кластеризации результатов поиска (1997) и другие. Все они отступили на задний план, когда в 1998 г. Л. Пейдж и С. Брин предложили алгоритм ранжирования PageRank, ставший изюминкой поисковика Google. В настоящее время Google обрабатывает большинство мировых поисковых запросов, а само это слово стало нарицательным понятием. 15 января

² InfoWorld, August 21, 1995, Vol. 17, Issue 34, p. 35

2006 г. глагол *google* был занесен в оксфордский словарь английского языка, а позже в словарь Merriam-Webster³. Кстати, о словарях: в 2001 году термин *рунет* появился в орфографическом словаре РАН В.В. Лопатина, а в 2005 году – и в орфографическом словаре Д.Э. Розенталя⁴.

В рунете создание навигационных инструментов началось несколько позже. В феврале 1996 года была впервые продемонстрирована поисковая машина Апорт (Е. Киреев). Тогда она искала только по сайту russia.agama.com. 8 октября 1996 года в интернете в свободном доступе появилась поисковая система Рамблер (Д. Крюков). Домен RAMBLER.RU зарегистрирован раньше, 26 сентября. 29 ноября того же года на семинаре РОЦИТ представлена первая версия каталога Ау! (А. Дыбенко и Ю. Поляк). 3 марта 1997 года открыт первый в рунете рейтинг-классификатор Rambler's TOP100, позволявший вести подсчет посещаемости сайтов. 23 сентября 1997 года на выставке Softool показана поисковая система Яндекс (А. Волож и И. Сегалович). 11 ноября 1997 года состоялась официальная презентация улучшенной версии машины Апорт. Разработчики этих систем регулярно встречались и поддерживали рабочие контакты.

"Русскоязычные поисковые системы в Интернет"

28 апреля 1998 года состоялся очередной семинар РОЦИТ. Заседание было посвящено теме "Русскоязычные поисковые системы в Интернет".

Семинар открыл председатель Правления РОЦИТ Андрей Зотов. Он рассказал о том, что Интернет в России - это уже свершившийся факт, и разработки различных компаний в сфере поиска информации - еще одно тому подтверждение. Развитие поисковых машин в русскоязычной части Интернет направлено на то, чтобы максимально облегчить пользователю решение многочисленных проблем и сделать удобным и комфортным поиск необходимой информации в Сети.

В первой части семинара выступили: Антон Никитин (компания "Intel"), Юрий Поляк (ЦЭМИ РАН), Валентин Куликов (компания "Ви-6"), Илья Сегалович (компания "CompTek"), Евгений Киреев (компания "Агама"), Дмитрий Крюков (компания "Stack Ltd."), Юрий Поляков (Информационное агентство "Интегрум-Техно").

Антон Никитин (Intel) рассказал о том, что Intel, как одна из ведущих компьютерных компаний мира, обычно не захватывает сферы бизнеса, отличные от основного направления. Однако, участие компании в развитии поисковых машин в России является серьезным маркетинговым шагом и позволяет не только еще раз показать имя компании и доказать возможности серьезно работать в области серверных приложений, но также вести важную статистику: откуда, когда и как люди ходят в Интернет. В настоящее время Intel поддерживает проекты "Апорт" и "Ау!" в русскоязычной части Интернет. Антон Никитин отметил, что компания-производитель процессоров считает одним из основных сдерживающих факторов для развития поисковых машин организацию дискового пространства серверов.

Юрий Поляк (ЦЭМИ РАН) представил Каталог-рубрикатор русскоязычных ресурсов Интернет "АУ!", который начал свою работу в Сети с ноября 1996 года. Каталог является совместным проектом Центрального экономико-математического института и компании Ви-6, занимающейся интеграцией Интернет технологий в России. Информационное наполнение полностью возложено на ЦЭМИ, а организацией сервера и сопровождением проекта занимается компания Ви-6. Юрий Поляк сообщил, что на данный момент в базе данных рубрикатора АУ содержится более 15 тысяч учетных записей, и она ежедневно пополняется. Он поблагодарил компанию Intel за сервер, который был предоставлен 27 апреля этого года специально под проект "АУ!"

Далее представитель компании Ви-6 и главный разработчик "АУ!" Валентин Куликов познакомил слушателей со строением каталога-рубрикатора.

Илья Сегалович (CompTek), рассказал об открытом полтора года назад проекте Яндекс. По его мнению Яндекс является идеальной поисковой машиной, предназначенной для поиска информации в текстах различной структуры с учетом морфологии русского и английского языка. Дружественный интерфейс Яндекс позволяет непрофессиональным пользователям вести поиск с помощью естественного языка. Яндекс - это полигон для испытания идей и одновременно локомотив, который толкает разработки компании в жизнь. Среди проблем, осложняющих работу поисковой машины Илья Сегалович назвал недостаточно формализованный стандарт русских кодеров, который осложняет задачу как разработчику, так и пользователю поисковой системы.

Евгений Киреев ("Агама") подробно рассказал о поисковой системе "Апорт", которая дает возможность пользователю осуществлять классический полнотекстовый поиск с учетом морфологии русского языка. Популярность "Апорта" определяется не рекламой, а числом запросов, которое неуклонно растет. Среди несомненных достоинств поисковой системы можно назвать возможность полнотекстовой реконструкции документа, если запрашиваемый сервер недоступен, поиск по датам, адресам, подзаголовкам, ссылкам и т.п., а также рейтинг наиболее популярных слов, используемых в запросах. Основной технической задачей поисковой системы Евгений Киреев назвал повышение производительности аппаратного и программного комплексов, а с точки зрения содержания - дальнейшая интеграция с Каталогом-рубрикатором "АУ!".

Дмитрий Крюков (Stack Ltd.) представил поисковую систему Rambler, которая заявила о себе в октябре 1996 года. В настоящее время в базе данных Rambler содержится более 3 млн. документов и

Семинар РОЦИТ 28.04.1998. Фрагмент пресс-релиза

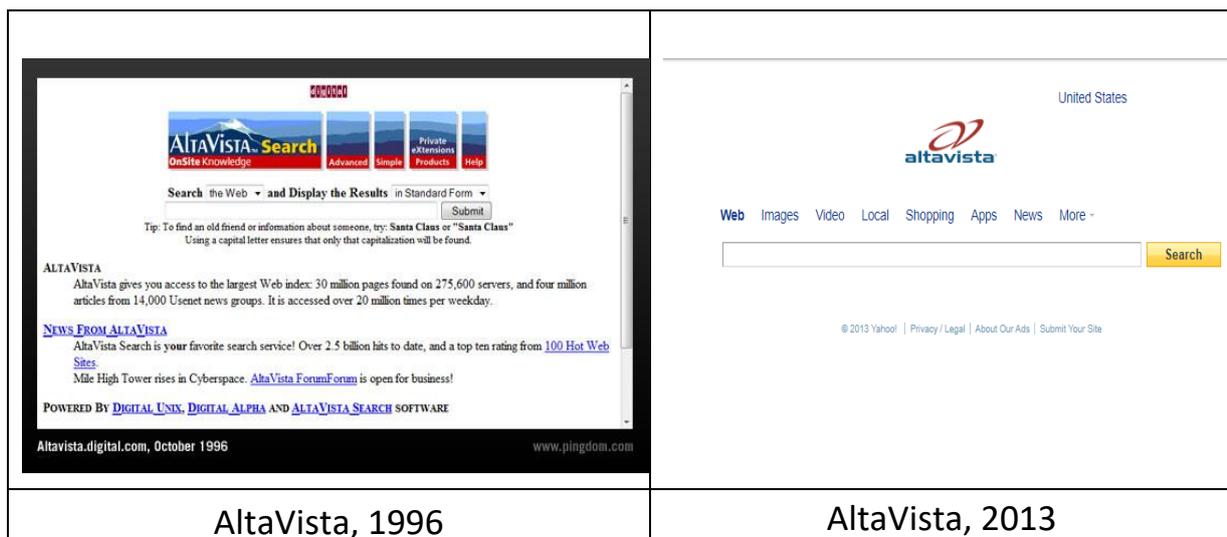
³ <http://wikireality.ru/wiki/Гуглить>

⁴ <https://computer.slovaronline.com/1088-RUNET>

Сейчас лишь немногие из этих систем остаются в рабочем состоянии, большинство остальных названий ничего не говорит широкой публике. В России без конкурентов остались Яндекс и Google. За океаном сохранились системы-ветераны WebCrawler и Yahoo! Следует, однако, сказать несколько слов о популярной в свое время системе AltaVista; в 1999 году ее доля в рунете среди всех сайтов составляла 16% (у Яндекса тогда было 14.5%)⁵. На момент появления AltaVista была самой быстрой из всех поисковых систем – она работала на мощнейшем на то время сервере DEC Alpha. Для автора она стала первым инструментом русскоязычного поиска, пусть довольно примитивным. В 1996 г. к ней присоединился ныне забытый HotBot, и только с появлением Яндекса поиск стал адекватно реагировать на запросы (Google такого уровня достиг лет через 10). А вот что сказал в одной из своих последних лекций сооснователь и первый технический директор Яндекса И.В. Сегалович: «Когда мы подключились к интернету, я обнаружил такую систему — AltaVista. Она закрылась три дня назад. В общем, наверное, никто не обратил внимания на эту новость, но люди, кто в интернете давно, — у них в этот момент скатилась скупая слеза ... когда мы вышли в интернет, и я посмотрел на AltaVista, мы поняли, что, вообще-то, можем сделать так же и даже лучше. Ощущение, что мы знаем как, и мы можем сделать лучше, — вот оно очень одухотворяло»⁶. Таким образом, AltaVista, кроме того, что стала первой обрабатывать русскоязычные запросы, еще и послужила источником вдохновения для Яндекса.

⁵ <https://www.kommersant.ru/doc/3417013>

⁶ <https://habr.com/company/yandex/blog/338562>. Примечание автора: AltaVista появилась 15 декабря 1995 г., закрылась 8 июля 2013 г. (<https://psychologyanswers.com/library/lecture/read/684527-is-altavista-dead>). Илье Сегаловичу к тому времени оставалось жить меньше трех недель.



ЯНДЕКС. ПЕРВЫЕ ГОДЫ

Итак, 23 сентября 1997 года посетители выставки Softool впервые смогли увидеть в действии новый инструмент поиска в интернете – поисковую систему Яндекс. На мой вопрос, чем она отличается от отечественных машин Апорт и Рамблер, американских разработок AltaVista и HotBot, главный редактор Яндекса Е. Колмановская пояснила, что поиск в системе производится на основе морфологического анализа текста, независимо от падежей и спряжений слов в запросе. Впервые реализован поиск на естественном языке для русского языка.

До регистрации одноименной компании оставалось три года, до IPO на бирже NASDAQ — почти 14 лет.

Яндекс появился не на пустом месте. К 1997 году в «портфеле» его команды уже был ряд серьёзных проектов: поиск по Международному классификатору изобретений (тогда ещё под управлением операционной системы MS DOS), электронный «Библейский компьютерный справочник» с поисковыми функциями, индексирование полного собрания сочинений А.С. Грибоедова по заказу Института мировой литературы РАН (уже на Windows). Для улучшения качества поиска, совершенствования системы морфологического распознавания слов в Академии наук была приобретена легальная электронная копия словаря русского языка. Грамматический словарь русского языка под редакцией академика А.А. Зализняка, итог многолетней работы, включает около 100 тысяч словоформ. Такая серьёзная научно-методологическая основа, дальнейшее сотрудничество разработчиков с группой академика Ю.Д. Апресяна из ИППИ РАН надолго предопределили

лидерство Яндекса в обработке русскоязычных запросов.

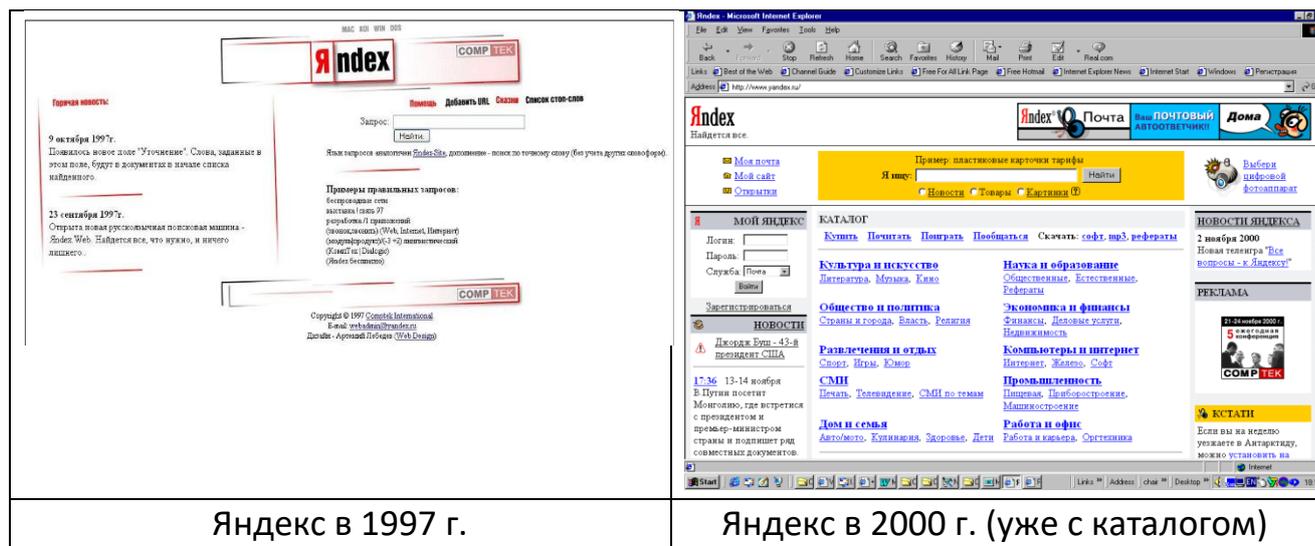


Сооснователи Яндекса Е. Колмановская, А. Волож (сидит) и И. Сегалович на Softool-1997. <https://fotostrana.ru/public/post/231942/2223059952>

Именно в эти годы были заложены важные принципы и решения, обеспечившие успешное развитие. Появились полезные для пользователей сервисы – Почта, Новости, Картинки, платёжная система Яндекс.Деньги, а также Яндекс.Словари, Яндекс.Маркет, хостинг сайтов. Запущена контекстная реклама, ставшая вскоре основной бизнес-моделью компании.

В 2000 г. зарегистрирована компания «Яндекс» (генеральный директор Аркадий Волож, директор по технологиям Илья Сегалович). Первым инвестором стал фонд ru-Net Holdings, вложивший \$5.28 млн. На предложения продать компанию руководители неизменно отвечали отказом.

Наша лаборатория сетевых информационных ресурсов ЦЭМИ РАН в 1999–2001 годах работала над концепцией каталога Яндекса. Совместно со специалистами Яндекса были разработаны структура и принципы классификации, затем выполнялись информационное наполнение базы данных, обучение редакторов. После ряда модернизаций каталог продолжал действовать до 2018 года. Кроме того, наш коллектив по заказу Яндекса сделал справочник по некоммерческим организациям.



Яндекс в 1997 г.

Яндекс в 2000 г. (уже с каталогом)

В 2001–2008 годах Яндекс проводил соревнования на Кубок по поиску в интернете, ставшие очень популярными [2]. По лицензии Яндекса подобные состязания проходили в России, Беларуси, Киргизии, Израиле, Норвегии и США. А завершилась «первая пятилетка» в 2002 году выходом Яндекса на самоокупаемость и первым проиндексированным терабайтом информации (это произошло 5 марта).

В последующие годы Яндекс стал в России безоговорочным лидером: в 2003 году он был признан самым посещаемым интернет-порталом России. Это произошло не только благодаря успехам компании, продолжавшей развивать алгоритмы поиска, но и вследствие провалов двух других поисковых машин (Апорт и Рамблер), сохранявших на рубеже веков примерный паритет. Появились и новые сервисы: Яндекс.Карты, Афиша, Лента, Словари. Для защиты от почтового спама на основе собственной технологии была запущена Спамоборона. Открылся поиск по блогам и форумам, но через некоторое время его пришлось закрыть – как сообщили, из-за накруток. Заработала Школа анализа данных — бесплатные двухгодичные курсы для подготовки специалистов, у которой затем появились филиалы в Новосибирске, Екатеринбурге, Минске, Киеве, а позже в Тель-Авиве и Беэр-Шеве. В Петербурге открылся первый удалённый офис разработки. И, наконец, был сделан первый шаг за пределы России: в 2005 г. компания открыла представительство в Одессе и стала развивать сервисы для украинской аудитории. В этом же году был запущен портал yandex.ua.

«Третья пятилетка» (2008–2012) стала, возможно, наиболее успешной. Ее

главный итог – первичное размещение акций (IPO) на Нью-Йоркской бирже NASDAQ⁷ 24 мая 2011 года. Акции продавались под тикером YNDX. 20 сентября того же года компания вышла за пределы постсоветского пространства: в Турции открыт портал yandex.com.tr. На нем были представлены Поиск, Карты, Почта и другие сервисы для турецких пользователей. Двумя годами раньше Яндекс начал работать в Казахстане и Белоруссии. На порталах yandex.kz и yandex.by представлены все основные сервисы (Почта, Карты, Авто, Расписания, Новости, Маркет), ориентированные на местную аудиторию, а поиск учитывает морфологию национальных языков. В 2008 году в Пало-Альто (Калифорния) основана Yandex Labs – центр научно-исследовательских разработок. Возглавил его Аркадий Борковский, который еще в 1990 г. вместе с А. Воложем основал компанию «Аркадия» и создал поисковый алгоритм, ставший прототипом Яндекса. В июне 2010 г. в Yandex Labs побывал тогдашний президент. В мае 2010 года Яндекс запустил англоязычную версию своей поисковой системы на домене YANDEX.COM.

	
<p>Посещение компании Yandex Labs. 24.06.2010. Калифорния, Пало-Альто. И. Сегалович, А. Волож, Д. Медведев, А. Борковский, В. Вексельберг. http://www.kremlin.ru/events/president/news/8152</p>	<p>Посещение офиса Яндекса 17.09.2017. http://static.kremlin.ru/media/events/photos/big2x/HJbyJzqrIHZEJf1r2mAoxaneJv38o359.jpg</p>

В этот же период стартовали такие продолжающиеся мероприятия, как технологическая конференция Yet another Conference (YaC-2010) и чемпионат по

⁷ NASDAQ (National Association of Securities Dealers Automated Quotation) - служба автоматизированных котировок Национальной ассоциации дилеров по ценным бумагам.
<https://www.nasdaq.com>

спортивному программированию Яндекс.Алгоритм (2011). Отметим такие оригинальные разработки, как новый метод машинного обучения Матрикснет, позволивший резко улучшить качество поиска; Яндекс.Браузер собственной разработки (за российским браузером последовали версии для Украины, Казахстана, Белоруссии и Турции); Яндекс.Диск – облачное хранилище данных, доступное на компьютерах, ноутбуках, смартфонах. Событием 2012 года стал запуск Навигатора с обновляемыми картами сотен городов, голосовым сопровождением и маршрутизацией с учетом пробок. Помимо Навигатора, за описываемый период появились мобильные приложения Поиск, Почта, Маркет, Музыка, Перевод, Директ, Яндекс.Деньги, Яндекс.Такси. По итогам 2012 года появилось сообщение, что Яндекс, обогнав Microsoft, стал четвертым поисковиком планеты (после Google, Baidu и Yahoo!) с 4.84 млрд поисковых запросов за месяц. При этом число поисковых запросов у Яндекса выросло за год на 28 %⁸

ЯНДЕКС. ЧЕРНАЯ ПОЛОСА

2013-й год оказался несчастливым: в июле не стало Ильи Сегаловича. Он был соучредителем и техническим директором Яндекса, создателем первой версии поисковой системы. Именно Илья придумал название компании. Его называли «душой Яндекса»⁹.

В течение следующих нескольких лет Яндекс сопровождали неприятности различного масштаба. Компания лишилась еще одного сооснователя: в ноябре 2012 года Елена Колмановская покинула пост главного редактора, а в середине 2016 года окончательно ушла из Яндекса. С её именем связывают PR-стратегию поисковика и создание образа Яндекса как интеллектуальной компании. Перед этим Лена неоднократно говорила о нарастающих трудностях в работе. Так, ей по должности приходилось отвечать на запросы от ФСБ по поводу счетов десятков пользователей, перечислявших деньги на те или иные проекты. Согласно законодательству, сервис обязан предоставлять информацию по запросу спецслужб. «Мы сами очень недовольны сложившейся ситуацией и разделяем негодование

⁸ https://www.vedomosti.ru/technology/articles/2013/02/07/yandeks_obognal_microsoft

⁹ <https://www.forbes.ru/forbes/issue/2013-09/243670-dusha-yandeksa>

пользователей», – говорила Колмановская. «После того, как наши юристы её рассмотрят, мы поймем, что можем сделать»¹⁰. Она также выслушивала многочисленные претензии по поводу якобы ангажированных результатов поиска и содержания новостей.

Здесь следует заметить, что в первые годы государство практически не обращало внимания на растущий интернет, и на редких встречах с начальством руководителей компаний на вопрос, чем им можно помочь, просили только не мешать. Безразличие стало постепенно сменяться тесными объятиями, порой удушающими. Стали появляться законы, регулирующие ИТ-отрасль – часто неграмотные и совершенно невыполнимые¹¹ (многие до сих пор вздрагивают, слыша название пьесы К. Тренёва «Любовь Яровая»).

Что касается Яндекс.Новостей, то они в какой-то момент (возможно, не по своей вине) перестали справляться со своей базовой задачей – рассказывать пользователю на главной странице о самых важных событиях. По данным компании, агрегатор ежедневно посещали больше 10 млн пользователей, и платформа часто становилась объектом критики с разных сторон. «В текущей политической ситуации существуют риски большого числа претензий от самых разных игроков, включая СМИ и политиков. Угодить враждующим акторам одновременно невозможно, и потребуются вводить цензуру. Цензура убыточна с точки зрения экономики, наносит урон имиджу холдинга»¹². А эмоциональный О.Ю. Тиньков высказался резко: «Яндекс – великая компания, Аркадий – великий предприниматель. Яндекс.Поиск – супергуд ... Аркадий большой молодец, но Яндекс.Новости – это то, за что он будет гореть в аду»¹³.

На подобные упреки Аркадий Волож ответил еще 31 мая 2017 г. в интервью корреспонденту «Ведомостей» А. Голицыной: «31 декабря 2016 г. у нас в сервисе было около 7000 источников, 1 января 2017-го их осталось примерно 1000. Только те, что имели лицензию Роскомнадзора на вещание»¹⁴. С тех пор законодатель-

¹⁰ <http://ru-compromat.livejournal.com/365585.html>

¹¹ kremlin.ru/news/20858, <https://tass.ru/ekonomika/5436971>, <https://lenta.ru/news/2018/07/31/yarovaya>

¹² <https://www.business-gazeta.ru/article/548496>

¹³ <https://vc.ru/offline/419530-ya-poteryal-9-mlrd-no-poschital-chto-pravilnee-umeret-s-horoshey-karmoy-glavnoe-iz-intervyu-olega-tinkova-dudyu>

¹⁴ <https://www.business-gazeta.ru/article/347528>

ство только ужесточалось, компания каждый раз подчинялась новым требованиям, и в итоге в топе Яндекса остались только «правильные», проверенные государством СМИ.

Вот ещё пример, как действия властей влияют на работу компании. После указа президента П. Порошенко от 15 мая 2017 г. о санкциях против юридических и физических лиц России¹⁵ стала невозможной работа киевского и одесского офисов, работу на Украине пришлось свернуть.

Дело не только в законах и указах. Случайно (или намеренно) брошенные высокопоставленными «специалистами» фразы о том, что интернет «используется для распространения детской порнографии, проституции, для доведения несовершеннолетних до самоубийств»¹⁶, что там «50 процентов порноматериалов»¹⁷ и «много информационного мусора, который очень часто подаётся как истина в последней инстанции»¹⁸, что «американцы пытаются монопольно управлять всемирной сетью»¹⁹ и т. п., способны обрушить капитализацию успешных ИТ-компаний²⁰ и перечеркнуть многолетний труд. Сталкивался с этим и Яндекс²¹. Даже «высочайший визит» накануне 20-летия компании не обошелся без шероховатостей²².

НОВЫЕ ДОСТИЖЕНИЯ

Несмотря на эти и другие проблемы, развитие продолжалось. Открылся офис разработки в Берлине, ставший 16-м офисом компании в мире и третьим в Западной Европе после Цюриха и Люцерна. На конференции YaC-2013 Яндекс представил собственную технологию распознавания речи SpeechKit и программ-

¹⁵ <https://www.rbc.ru/politics/16/05/2017/591aae3e9a79473f81fb65b7>

¹⁶ Расширенное заседание коллегии МВД России 3 марта 2021 года.
<http://www.kremlin.ru/events/president/transcripts/65090>

¹⁷ <http://kremlin.ru/events/president/transcripts/6693>

¹⁸ <http://www.kremlin.ru/events/president/news/66554>

¹⁹ <https://raec.ru/live/branch/7072/>

²⁰ <http://www.finmarket.ru/news/896192>

²¹ <https://rusevik.com/politika/66412-putin-obrushil-yandeks.html>,
<https://www.rbc.ru/business/11/10/2019/5da08b679a7947ba554d8583>

²² https://ru.wikinews.org/wiki/Путин_парализовал_Яндекс,
<http://kremlin.ru/events/president/news/55674>

ный интерфейс для работы с ней. А через год заработали облачный сервис распознавания речи и приложение, которое записывает текст под диктовку и озвучивает написанное. Появился сервис для совместной работы внутри организации — Яндекс.Коннект. Начала работать Yandex Data Factory — направление по работе с «большими данными» для компаний, обрабатывающих большие массивы информации. В нем использовались технологии Яндекса — машинное обучение, распознавание образов и речи, нейронные сети, обработка естественного языка.

Запущен Яндекс.Store — магазин приложений для Android. В 2015 году появились мобильное приложение для водителей Яндекс.Парковки и музыкальный сервис Яндекс.Радио, где можно найти музыку разных жанров и эпох. Заработал Яндекс.Транспорт — мобильный сервис, который показывает, где находится нужный автобус или трамвай и когда он приедет на остановку. Запущено мобильное приложение для записи к врачу Яндекс.Здоровье.

В 2014 г. Яндекс открыл в НИУ «Высшая школа экономики» факультет компьютерных наук. Он готовит специалистов по направлениям «Прикладная математика и информатика» и «Программная инженерия». Образовательную деятельность продолжил проект Яндекс.Лицей. Это очная программа обучения для школьников 8–9 классов, которые хотят получить навыки программирования. В 2016 году проект начал работать в Калуге, Пензе, Саратове и Тамбове. К настоящему времени магистерские программы в партнёрстве с Яндексом работают в МФТИ, университете ИТМО, Уральском федеральном университете, Нижегородском университете им. Н.И. Лобачевского, Европейском университете в Санкт-Петербурге, Белорусском госуниверситете.

Кроме того, Яндекс вплотную занялся вопросами искусственного интеллекта. Он создал голосового помощника Алису на основе нейронной сети. Она помогает в повседневных делах, обращаясь к сервисам Яндекса: может включить музыку, рассказать о погоде или построить маршрут. В 2017 г. компания провела испытания беспилотных автомобилей. В них участвовали две машины, оснащённые датчиками для сбора данных и программным обеспечением. Эти прототипы самостоятельно передвигались по заданному маршруту, распознавали и объезжали препятствия. В декабре 2023 стало известно, что Яндекс вошел в число мировых лидеров в области развития искусственного интеллекта согласно рейтингу,

опубликованному учеными из Массачусетского технологического института²³.

Много новостей принесли и последующие годы. Постоянно продолжали совершенствоваться поисковые алгоритмы, каждый год появлялись новые решения. Но Яндекс уже давно не столько поисковик, сколько экосистема, где вырастают новые сервисы и бизнесы, подпитывающие весь организм и финансово, и технологически. Число специализированных сервисов превышает 120, они ежедневно нужны пользователям и значительно упрощают жизнь. Практически каждый продукт уникален. В этот период появились сервисы, облегчающие жизнь горожан, это каршеринг Яндекс.Драйв, экспресс-доставка продуктов и товаров для дома Яндекс.Лавка и сервис по доставке еды Яндекс.Еда с роботом-курьером Яндекс.Ровер. Они оказались особенно полезны из-за режима самоизоляции во время пандемии коронавируса, когда люди учились жить в условиях карантина. Заказы еды, продуктов, такси, доставку посылок, аренду машин объединило приложение Яндекс.GO.

Яндекс активно скупает и поглощает маленькие ИТ-компании и стартапы, среди которых платформы по управлению интернет-рекламой, сервис по продвижению товаров, студия разработки мобильных приложений, стартап по распознаванию рукописного и печатного текста. После того как в 2020 году владельцем сервиса Яндекс.Деньги стал Сбербанк, Яндекс купил банк «Акрополь», получил банковскую лицензию и открыл Яндекс.Банк. В 2022 г. была запущена виртуальная карта «Мир».

Начиная с 2018 года, в продажу выходит линейка «умных колонок». Эти беспроводные bluetooth-устройства собственной разработки с голосовым помощником Алисой воспроизводят музыку в высоком качестве, показывают на телеэкране видео, фильмы и сериалы, рассказывают детям сказки, помогают в повседневных делах. В ноябре 2022 г. выпущен умный телевизор от Яндекса с той же Алисой для голосового управления.

Совершенствуются традиционные сервисы. Яндекс.Переводчик внедрил гибридную систему перевода с использованием технологии нейронных сетей. На Яндекс.Погоде появились красивые погодные карты всей планеты. Открыты две

²³ <https://www.forbes.ru/tekhnologii/502105-ucenye-iz-mit-i-eksperty-vklucili-andeks-v-cislo-mirovyh-liderov-po-razvitiu-ii>

тестовые зоны (в Иннополисе и Сколково), где можно прокатиться на машине без водителя. Суперкомпьютер Яндекса «Червоненкис» (названный в честь А.Я. Червоненкиса, одного из крупнейших теоретиков машинного обучения) вошёл в двадцатку самых мощных в мире и стал самым производительным в Восточной Европе.

ЖИЗНЬ В НОВЫХ УСЛОВИЯХ. 2022–2023

Свой юбилейный, 25-й год Яндекс встретил в статусе крупнейшей российской ИТ-компании с отделениями и офисами во многих странах. На базе поиска создан разветвленный диверсифицированный бизнес, образовавший, как принято говорить, мощную экосистему с миллионами пользователей. В начале 2022 года у Яндекса были большие планы на будущее. Его американскую «дочку» по производству беспилотных автомобилей аналитики Morgan Stanley называли одним из лидеров на мировом рынке и оценили в \$7 млрд. Дроны Яндекса по соглашению с сервисом доставки Grubhub развозили еду по кампусам университетов Огайо и Аризоны. Другой международный бизнес Яндекса, подразделение по разработке системы управления базами данных ClickHouse было выделено в отдельный бизнес в США и привлекло \$300 млн инвестиций при оценке \$2 млрд. ClickHouse использовали Uber, Tesla, Spotify, Bloomberg, Alibaba и ByteDance.

Прошедший год внезапно и необратимо изменил жизнь страны и каждого из нас. Это не обошло стороной и Яндекс, поступательное развитие которого было резко прервано. В результате санкций заметно сократилось число зарубежных офисов и филиалов, которые Яндекс создавал на протяжении многих лет. Многие связи оказались замороженными, а часть зарубежных бизнесов – парализованной. Так, летом 2022 г. прекратили работу сервисы экспресс-доставки продуктов от Яндекса в Париже и Лондоне, открытые по модели Яндекс.Лавки соответственно в августе и октябре 2021 г. (первым таким заграничным подразделением стал в ноябре 2020 г. филиал в Тель-Авиве, который продолжает работу).

На государственном уровне применили санкции страны Балтии: Латвия и Эстония запретили деятельность Яндекса как оператора такси, в Литве «Яндекс Go» был убран из местных разделов магазинов приложений. Про украинские офисы сказано выше. В Финляндии местные энергетики отключили дата-центр Яндекса от электропитания, из-за чего пришлось перейти на работу от дизельных

генераторов. Напомним, этот дата-центр в городке Мянтсяля теплом, выделяемым серверами, способствует отоплению города²⁴ [3].

После 24 февраля 2022 года Яндекс отозвал своих роботов – доставщиков еды из американских университетов. Тестирование беспилотных автомобилей Яндекса в штате Мичиган также остановлено. После решения Grubhub прекратить сотрудничество от партнерства с российской ИТ-компанией отказался американский поисковик DuckDuckGo, а разработчики браузера Mozilla Firefox исключили Яндекс из списка стандартных поисковиков²⁵.

Утрата одних бизнес-направлений активизировала поиски других. В этом плане перспективными представляются страны Латинской Америки – Эквадор, Бразилия, Чили, Перу, Мексика. Яндекс нанимает сотрудников для работы над доставкой в ЮАР и Турции, а также планирует развивать запущенный в марте проект в Сербии. Сервисы доставки и прогноза погоды работают под названиями, явным образом не связанными с материнским брендом Яндекса: первый использует вариацию бренда YanGo, запущенного в 2018 году в Кот-д`Ивуаре, второй — Meteum²⁶. В декабре появилось сообщение, что Яндекс создал международный бизнес с десятком стартапов. С этой целью компания зарегистрировала в Армении и США новую холдинговую компанию Beyond ML, чтобы избежать проблем на западных рынках на фоне санкций²⁷.

В список стран, где есть подразделения Яндекса, входят Армения, Белоруссия, Израиль, Казахстан, Киргизия, Китай, Нидерланды, Сербия, США, Турция, Узбекистан, Чехия, Швейцария. Офис в Белграде становится основным для Европы²⁸. Офисы Яндекса работают в 20 российских городах²⁹.

На бирже NASDAQ с 28 февраля 2022 года были приостановлены торги акциями Yandex, а 15 марта 2023 года Яндекс получил уведомление о делистинге,

²⁴ <http://www.rbc.ru/rbcfreenews/56975d719a7947a05b2499a6>

²⁵ <https://www.forbes.ru/tekhnologii/483281-razdevalka-ekosistemy-cto-proishodilo-s-andeksom-v-2022-godu>

²⁶ <https://www.kommersant.ru/doc/5482305>

²⁷ <https://www.kommersant.ru/doc/5736027>

²⁸ <https://www.forbes.ru/tekhnologii/502786-andeks-pereregistroval-prilozhenia-v-app-store-i-google-play-na-serbskie-urlica>

²⁹ <https://yandex.ru/jobs/locations>

то есть исключения акций компании из оборота на бирже. Яндекс направил апелляцию, и 8 июня она была удовлетворена: NASDAQ отменил это решение, оставив в силе приостановку торгов. На московской бирже акции Яндекса продолжают котироваться³⁰.

Последние события оказали огромное негативное влияние на судьбы людей, вложивших массу энергии и знаний в становление и развитие компании. Об этом, в частности, пишет издание The Bell в материале «История человеческих трагедий»³¹. Из совета директоров Яндекса вынуждены были уйти предприниматель и общественный деятель Эстер Дайсон [4] и профессор Стэнфордского университета Илья Стребулаев³². 3 июня 2022 г. в санкционный список попал основатель и крупнейший акционер Яндекса А. Волож³³. Ему, в частности, вменяется в вину, что он «является ведущим бизнесменом, работающим в секторах экономики, обеспечивающих существенный источник дохода правительству Российской Федерации ... Кроме того, как основатель и генеральный директор Яндекса он оказывает материальную или финансовую поддержку правительству Российской Федерации и несет ответственность за поддержку действий или политики, которые подрывают или угрожают территориальной целостности, суверенитету и независимости Украины». В личном комментарии Волож назвал это решение Еврокомиссии нелогичным, ошибочным и контрпродуктивным, однако сразу после публикации списка он покинул все свои посты в компании, а право голосования своим пакетом (45.6% голосов) передал совету директоров. Ранее (15.03.2022) под санкции попал исполнительный директор Яндекса Т. Худавердян³⁴ за то, что 24 февраля он был в Кремле на встрече бизнесменов с президентом. Заметим, что в отличие от многих российских медиа, компаний и банков, сам Яндекс избежал официальных санкций: ни материнская компания Yandex N.V. (Нидерланды), ни какие-либо компании группы не находятся под санкциями США, Евросоюза, Швейцарии или Великобритании.

³⁰ https://yandex.ru/company/press_releases/2023/21-03-23

³¹ <https://thebell.io/istoriya-chelovecheskikh-tragediy-kak-voyna-razvela-arkadiya-volozha-s-menedzherami-yandeksa>

³² <https://www.kommersant.ru/doc/5249698>

³³ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2022:153:FULL>

³⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022D0429>

После антивоенного заявления А.Ю. Воложа от 10.08.2023³⁵ была сделана попытка обжаловать решение Еврокомиссии, но успеха она пока не имела³⁶.

Так или иначе, но под угрозой санкций команда была вынуждена спешно (и, скорее всего, запоздало) избавляться от самых токсичных активов - Яндекс.Новостей, а также рекомендательной платформы Яндекс.Дзен. Стратегически от продажи «Новостей» и «Дзена» зависела сама возможность дальнейшего существования Яндекса. В результате комбинации с участием VK и Сбербанка Яндексу достался сервис доставки еды Delivery Club в обмен на «Новости» и «Дзен». Также Яндекс отдал VK свою титульную страницу (домен YANDEX.RU), и теперь вход в поиск Яндекса возможен только по адресу <https://ya.ru>.

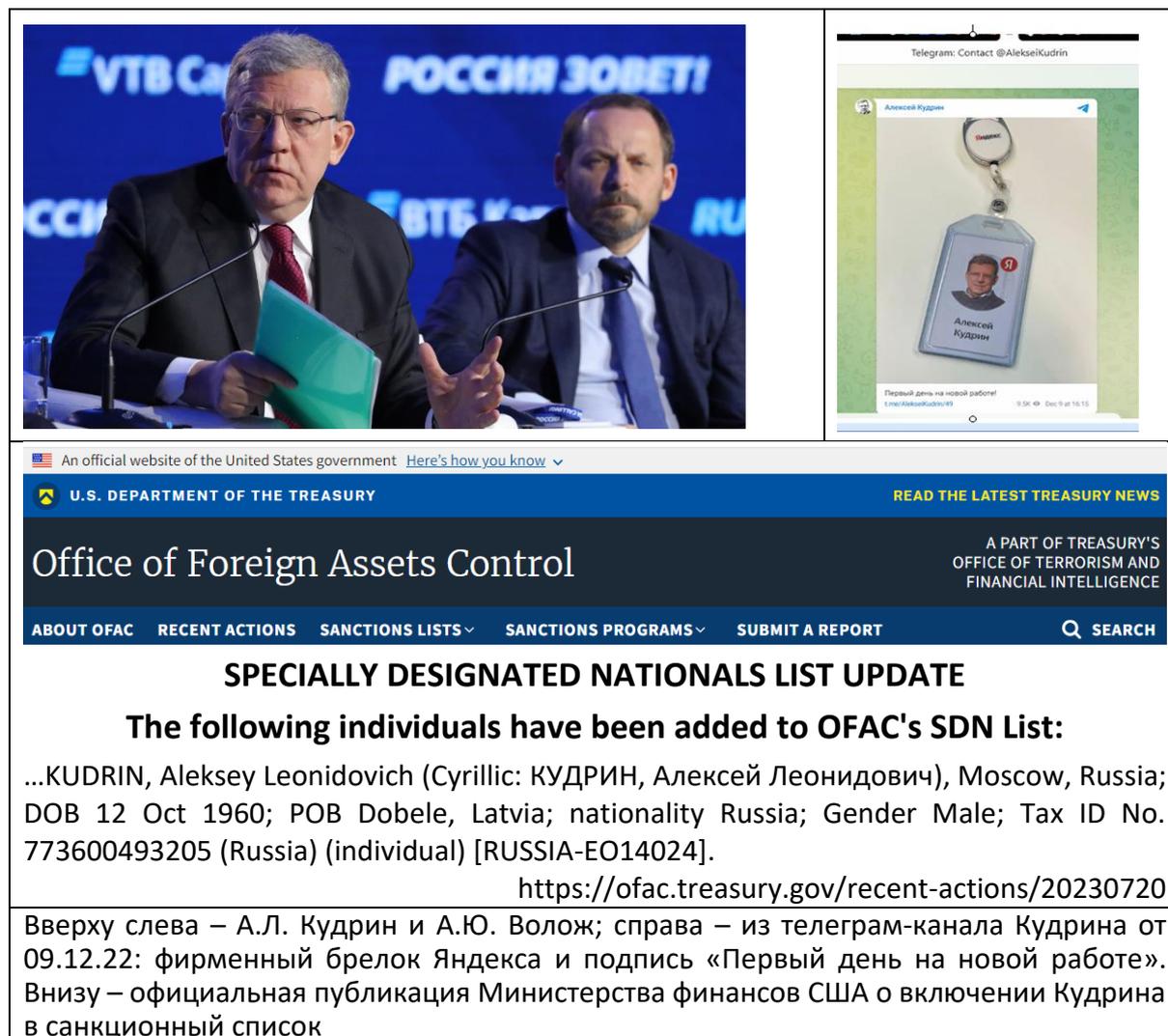
25 ноября 2022 г. совет директоров Yandex N.V. объявил о намерениях внести изменения в структуру собственности и корпоративного управления. Речь идет о выделении основной части бизнеса, включая ряд зарубежных направлений, в отдельную группу компаний, которая сохранит за собой бренд Яндекса. При этом Yandex N.V. планирует со временем выйти из числа акционеров этой группы компаний и сменить своё название. Важная роль в осуществлении этих реформ отводится А.Л. Кудрину, который в связи с этим ушел из Счетной палаты и занял в Яндексе пост советника по корпоративному развитию³⁷. Крупнейшим голосующим акционером нового Яндекса должен стать «коллективный Волож» – специальный фонд под управлением Кудрина и трех топ-менеджеров. Тем временем в отношении Кудрина 20.07.2023 введены санкции Минфина США³⁸.

³⁵ https://t.me/thebell_io/24062

³⁶ <https://www.bloomberg.com/news/articles/2023-12-13/yandex-founder-calls-for-sanctions-removal-after-criticism-of-russian-invasion>

³⁷ <https://www.forbes.ru/tekhnologii/482300-aleksej-kudrin-vysel-na-rabotu-v-andeks>

³⁸ <https://ofac.treasury.gov/recent-actions/20230720>



Вверху слева – А.Л. Кудрин и А.Ю. Волож; справа – из телеграм-канала Кудрина от 09.12.22: фирменный брелок Яндекса и подпись «Первый день на новой работе». Внизу – официальная публикация Министерства финансов США о включении Кудрина в санкционный список

Дискуссии о схеме раздела Яндекса должны закончиться к концу 2023 года. Однако на середину декабря ясности нет. Планировалось, что потенциальные инвесторы выкупают экономическую долю в компании, а голосующий контроль останется за командой Яндекса в форме фонда. Затем возник вариант с передачей «реального контроля» общей структуре, представляющей интересы «тщательно спрятанной» группы инвесторов. «Голосующая доля у общей структуры будет де-факто контролирующая. С некоторыми правами у менеджеров компании и Кудрина по ограниченному кругу вопросов», — сообщает Forbes³⁹. Среди потенциальных инвесторов в российский бизнес Яндекса нет единого мнения по вопросу обмена акций, происходит борьба между новыми инвесторами. «Сделку

³⁹ <https://www.forbes.ru/tekhnologii/500982-investory-obsuzdaut-pokupku-andeksa-cerez-edinuu-strukturu>

осложняет то, что Яндекс должен одновременно удовлетворять требованиям российских властей, которые не приветствуют сделки с большим выводом капитала из страны и «выпуском» замороженных иностранных инвесторов, но и не может поступиться интересами зарубежных акционеров»⁴⁰.

Что касается А.Ю. Воложа, он был намерен под управлением Yandex N.V. развивать за рубежом беспилотники, облачные сервисы, «Яндекс.Практикум» и «Толоку» – краудсорсинговую площадку для тестирования и улучшения системы с помощью искусственного интеллекта⁴¹. 30 декабря 2022 г. Волож написал прощальное сообщение для сотрудников Яндекса⁴².

27 октября 2023 года вышел последний к настоящему моменту подробный пресс-релиз Яндекса об итогах III квартала 2023 года [5]. Как следует из отчёта, доля компании на российском поисковом рынке (включая поиск на мобильных устройствах) в III квартале составила в среднем 62.6%. Численность сотрудников достигла 25.7 тысяч человек. В отчете за II квартал говорится о 24.3 тыс. чел. Это дает представление о динамике роста. Вспоминается прошлый век, тесное помещение на улице Губкина в пристройке к Институту общей генетики и 10–15 исключительно талантливых сотрудников – компания ComТек, будущий Яндекс (Аркадий говорил тогда, что принимает на работу только гениальных людей – тогда так и было, о нынешних тысячах вряд ли это можно сказать). Коллектив быстро рос, при регистрации компании «Яндекс» команда насчитывала 25 человек, а в 2001 г. после переезда в первый собственный офис Яндекса в ВЦ РАН (в зал, где ранее стояла машина БЭСМ, а позже Эльбрус) – уже 60. Через 5 лет возникла необходимость в новом расширении, и Яндекс переехал в бывшее здание ткацкой фабрики В.Е. Филиппова на Самокатной улице. Архитектурное бюро «Атриум» в 2004–2005 годах вписало в кирпичные стены XIX века современный хайтековский эргономичный офис с потрясающей атмосферой и «живыми изгородями» – стеклопакетами с живой травой внутри. Он был одинаково удобен и для разработок, и для досуга, и для совещаний. Яндекс располагался здесь до 2010 года.

⁴⁰ <https://www.forbes.ru/tekhnologii/502456-u-investorov-voznikli-raznoglasiya-po-sdelke-s-andeksom>

⁴¹ <https://toloka.ai>

⁴² <https://www.forbes.ru/tekhnologii/483281-razdevalka-ekosistemy-cto-proishodilo-s-andeksom-v-2022-godu>



Здесь жил Яндекс. Вверху – первый офис в ВЦ РАН на ул.Вавилова, 40 (<https://yandex.ru/blog/company/dom-kotoryy-postroit-yandeks>) и здание на ул. Самокатной 1, стр 21 (<https://yandex.ru/company/offices/moscow/samokatnaya>); внизу – штаб-квартира на ул. Льва Толстого, 16 (<https://ru.wikipedia.org/wiki/Яндекс>) и стройка нового здания на улице Косыгина, 15 (Академика Зельдовича, 1) осенью 2023 г. – фото автора

Но продолжающийся рост компании потребовал нового расширения площадей, пришлось отказаться от этого замечательного здания. Новой штаб-квартирой Яндекса стал бизнес-центр «Красная Роза» – квартал между улицами Льва Толстого и Тимура Фрунзе на месте территории одноименной крупнейшей шелкоткацкой фабрики. Общая площадь всех зданий — более 50 тысяч квадратных метров. Переезд был весьма своевременным: уже к 2012 году количество сотрудников насчитывало 4.5 тысяч человек, а серверный парк – 10 тысяч машин; объём

рынка интернет-рекламы превысил \$1.7 миллиардов. Рубеж в 5 тысяч сотрудников Яндекс преодолел в 2014 году; в 10 тысяч – в 2019-м⁴³.

Все эти годы продолжалось расширение за счет аренды в московских бизнес-центрах. В 2017-м открылся офис на Садовнической улице в комплексе «Аврора бизнес парк» площадью 33000 кв. м. В 2020 году Яндекс.Маркет переехал в БЦ Lotte Plaza и занял пять этажей (15800 кв. м). Открылись два офиса на территории «Москва-Сити» — в БЦ «ОКО» (6 этажей) и «ОКО II» (9 этажей). В 2021 г. арендованы 20833 кв. м в комплексе Neva Towers. Но и этого не хватает.

В настоящее время подходит к концу строительство нового комплекса между улицами Академика Зелинского и Академика Зельдовича площадью более 262 000 квадратных метров⁴⁴. Здание имеет вид трех крыльев высотой от 12 до 15 этажей, объединенных центральным атриумом. Кроме того, предусматривается еще пять подземных этажей.

Пока идут дискуссии о будущем Яндекса, он продолжает оставаться самой дорогой компанией рунета. В 2016 году его капитализация составляла \$7.6 млрд (по оценке Forbes) [3]. Это лидерство сохранилось до настоящего времени. За последний год капитализация снизилась в 1.7 раза (с \$17.4 млрд в 2022 г. до \$10.2 млрд в 2023 г.), но Яндекс удержал первенство (на 2-м месте Wildberries с \$9.8 млрд, на 3-м – Ozon с \$5.2 млрд)⁴⁵. Среди всех российских компаний Яндекс (по состоянию на 4 августа 2023 г.) с капитализацией 0.985 трлн руб. лишь 14-й, а впереди Сбер (6.04 трлн), Роснефть (5.58) и Новатэк (4.84). Газпром, многолетний лидер, по понятным причинам опустился на 5 место⁴⁶. Интересно, что в мире ИТ-компании лидируют. Первые два места занимают Apple (\$3.072 трлн) и Microsoft (\$2.755 трлн), в пятерку также входят Alphabet (Google) и Amazon. В этом списке Walmart – 17-я, Mastercard – 20-я, Nestlé – 26-я. На 35-м месте Bank of America, на 48-м – McDonald, на 98-м – General Electric⁴⁷.

⁴³ <https://smart-lab.ru/q/YNDX/MSFO/employees>

⁴⁴ <https://rg.ru/2023/05/31/iandeks-gde-najdetsia-vsyo.html>

⁴⁵ <https://www.forbes.ru/tekhnologii/485182-30-samyh-dorogih-kompanij-runeta-2023-rejting-forbes>

⁴⁶ <https://www.tinkoff.ru/invest/social/profile/AK47pfl/c38fabe1-f0a5-4161-bdb4-09af44918f16/>

⁴⁷ <https://companiesmarketcap.com>

GOOGLE

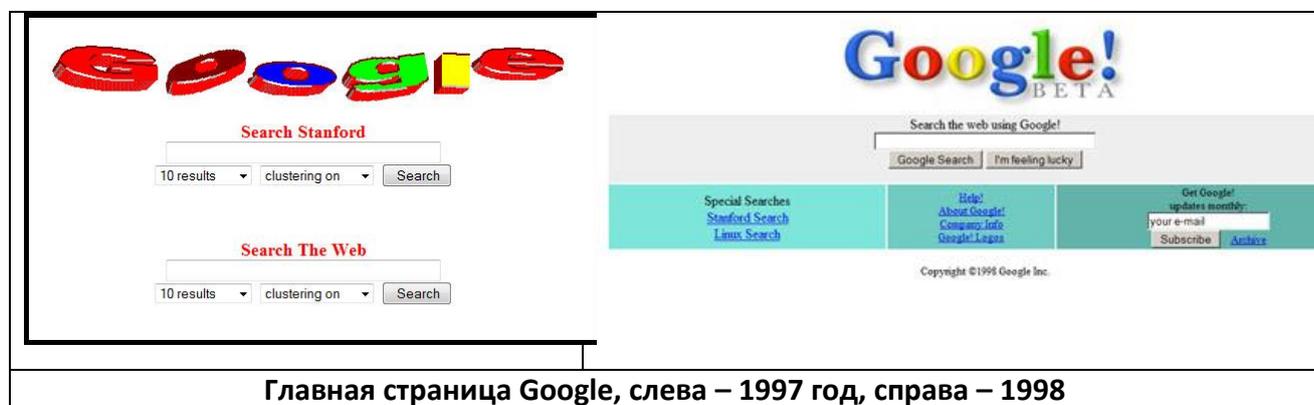
В отличие от Яндекса, ведущая мировая поисковая система развивается без турбулентностей; ее капитализация в ноябре 2021 года достигла 2 триллионов долларов. Как и Яндекс, она недавно отметила 25-летний юбилей: домен GOOGLE.COM был зарегистрирован 15 сентября 1997 года (изначально поисковая система находилась на сайте Стэнфорда google.stanford.edu); компания Google была официально зарегистрирована 4 сентября 1998 г.

У системы прослеживаются «русские корни». Сергей Брин родился в Москве 21 августа 1973 года и просто не мог не стать математиком. Его мама, Женя Краснокутская (однокурсница автора), окончила механико-математический факультет МГУ в 1971 году; отец, Михаил Брин, – годом раньше. Дед, Израиль Абрамович Брин (1919–2011), был доцентом кафедры высшей математики МЭИ. В октябре 1979 года семья переехала в Америку, где Михаил стал профессором Мэрилендского университета, Евгения работала в NASA, а Сергей, получив степень бакалавра в Мэриленде, поступил в магистратуру Стэнфордского университета. Там в 1995 году он познакомился с Ларри Пейджем.

Лоуренс Эдвард Пейдж (Ларри) родился 26 марта 1973 г. в семье сотрудников Мичиганского университета: отец – профессор информатики, мать – преподаватель программирования. С самого раннего возраста он интересовался персональными компьютерами. После школы он стал бакалавром по компьютерной инженерии в Мичиганском университете и магистром по информатике в Стэнфорде.

Совместная работа Брина и Пейджа «Анатомия системы крупномасштабного гипертекстового интернет-поиска» легла в основу проекта, который сейчас весь мир знает как Google. Авторы предложили новый принцип обработки данных при поиске информации в глобальной сети. В его основе лежала идея о том, что важность отдельной веб-страницы тем выше, чем больше других страниц на нее ссылаются. Предложенный алгоритм ранжирования страниц получил название PageRank, и вскоре выяснилось, что основанная на нем поисковая система превосходит существовавшие тогда аналоги, которые оценивали страницы по количеству появления на них искомого слова. Заметим, что к этому времени Робин

Ли уже создал небольшую поисковую систему RankDex (1996) на базе аналогичной стратегии для ранжирования страниц и впоследствии использовал запатентованную технологию RankDex при создании Baidu в Китае.



Первое финансирование будущая компания получила в августе 1998 года: соучредитель Sun Microsystems Андреас Бехтольсхайм после демонстрации возможностей поисковика выписал чек на 100 000 долларов. На эти деньги была основана компания Google Inc. А первым офисом стал гараж на окраине Менло-Парка, штат Калифорния. Его арендовали у Сьюзан Воджиски (впоследствии она стала генеральным директором YouTube). Два замечания: через несколько лет Брин женится на ее младшей сестре Анне (Энн), а польская фамилия Wojcicki также читается как Войчицки.

К концу 1998 года Google проиндексировал 60 миллионов страниц. Рос штат компании, вскоре в гараже стало тесно, и в марте 1999 года компания переехала в город Пало-Альто. Но по мере расширения бизнеса к лету 2003 года пришлось искать более просторное рабочее пространство, чтобы разместить 1000 с лишним сотрудников. Был арендован (а затем приобретен) комплекс зданий в Маунтин-Вью – бывшее помещение Silicon Graphics, – где и остаётся сегодня. Этот гигантский комплекс получил название Googleplex. Его площадь после дополнительного строительства составляет 190 тысяч квадратных метров, адрес: 1600 Amphitheatre Parkway, Mountain View, CA 94043. В 2010 году компания приобрела более вместительное здание в Нью-Йорке по адресу 111 Eighth Avenue.



Googleplex

Первый зарубежный филиал Google был открыт в 2001 году. Сейчас в 40 с лишним странах работает более 70 представительств.

В марте 2001 года к команде присоединился опытный управленец Эрик Шмидт, бывший генеральный директор Novell. Благодаря своему опыту и таланту, он вскоре вошел в совет директоров, а в августе занял должность генерального директора Google. Пейдж при этом стал называться президентом по продукции, а Брин – президентом по технологии. Пресса писала, что Шмидт «стал единственным взрослым на детской площадке». Чтобы творить и не отвлекаться, молодым гениям был необходим тот, кто возьмет на себя технические, кадровые и организационные вопросы⁴⁸. Пост генерального директора Шмидт занимал до апреля 2011 года. После этого до конца 2017 года он работал в качестве исполнительного председателя компании.

Первичное размещение акций на NASDAQ состоялось 19 августа 2004 г. В апреле того же года публике был представлен Gmail. Этот почтовый сервер предоставлял расширенный поиск, 1 Гб дискового пространства и ряд других функций.

⁴⁸ <https://dzen.ru/a/YI3icnMtdyqgcqrb>



Сверху вниз: Л. Пейдж, Э. Шмидт, С. Брин. Time, 20 февраля 2006 г.
<https://content.time.com/time/magazine/0,9263,7601060220,00.htm>

В 2005 году Google представил картографическую систему Google Maps (позднее, в 2009 г., заработала навигация GPS для мобильных телефонов).

В 2006 году Google купил видеохостинг YouTube за \$1.65 млрд.

В сентябре 2008 года появился браузер Google Chrome, который в последующие четыре года стал самым популярным веб-браузером для смартфонов и компьютеров в мире.

В том же месяце компания выпустила T-Mobile G1, первый телефон на базе ОС Android, купленной в сентябре 2005 года за \$50 миллионов.

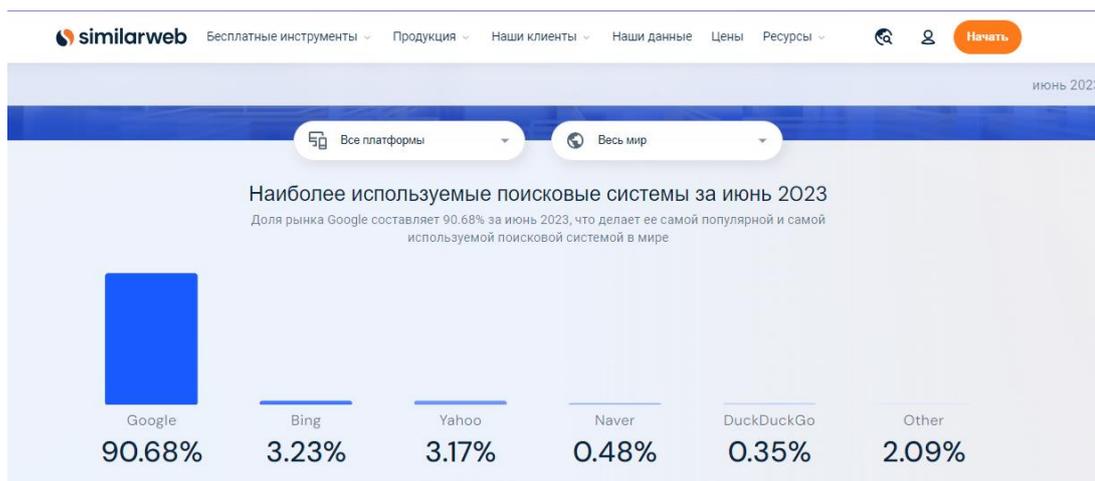
В 2010 году Брин организовал и возглавил дочернюю компанию Google X для разработки беспилотных автомобилей, VR-очков Glass и других технологических новинок.

По статистике около 82% пользователей интернета ищут информацию через поисковик Google, а весь рынок мобильных ОС на 85% состоит из систем Android⁴⁹.

Кроме того, Google предоставляет облачное хранилище данных Google Диск; магазин приложений Google Play (Play Market); Google Translate выполняет перевод слов, фраз, текстов и целых страниц.

С огромным отрывом Google является поисковиком №1 в мире и почти во всех странах (о немногих исключениях см. в заключительном разделе).

⁴⁹ <https://gb.ru/blog/kompaniya-google>



<https://www.similarweb.com/ru/engines/>

ЯНДЕКС VS GOOGLE

И еще немного статистики⁵⁰.

	Яндекс	Google
Дата основания	23 сентября 1997 года	4 сентября 1998 года
Выручка за 2022 год	521.7 млрд рублей	279.8 млрд долларов
Сотрудники	25.7 тыс человек	139 тыс человек
Визиты в поиске (Россия)	1.12 млрд/месяц (63.56%)	623 млн/месяц (35.78%)
Рейтинг сайта в России	1-е место	2-е место
Рейтинг сайта в мире	9-е место	1-е место
Браузер в России, доля	41.81 % (Яндекс Браузер)	29.70 % (Google Chrome)
Посещаемость сайта	3.3 млрд визитов/месяц	87 млрд визитов/месяц

А теперь вспомним то уже отдаленное время, когда Google был еще в стадии становления и готовился к экспансии на развивающиеся рынки. Россия в этом плане выглядела заманчиво, но американский поисковик не владел морфологией русского языка. Зато это было главной «фишкой» быстро растущего Яндекса. Google заинтересовался российской ИТ-компанией, Брин и Пейдж приехали в московский офис на переговоры о возможной сделке. Разработчики и менеджеры Яндекса с энтузиазмом восприняли интерес американцев, мечтая об общем офисе с коллегами из Google в Silicon Valley. Но довольно скоро выяснилось,

⁵⁰ <https://inclient.ru/yandex-stats>

что речь идет, как в сказке о Красной Шапочке, не столько о слиянии с серым волком, сколько о поглощении. Равноценного сотрудничества не получилось, Яндекс от предложения отказался. Однако через некоторое время (точнее, в 2004 году) Google самостоятельно пришёл в Россию и начал процент за процентом стремиться к лидерству. По данным службы WHOIS, домен GOOGLE.RU зарегистрирован 04.03.2004⁵¹ (для справки: домен YANDEX.RU – 23.09.1997).

Однако регистрация домена – лишь малая часть дела; поисковик необходимо постоянно совершенствовать. В первые годы поисковые системы порой выдавали странные и даже анекдотичные результаты. Прочитируем статью технического директора Яндекса И.В. Сегаловича. «Широко известна история, когда по слову *геморрой* Яндекс десятой ссылкой выдавал *microsoft.com* (этот же сайт выдавался на запросы *жопа* и *мастдай*). Связано это с тем, что Яндекс учитывает текст ссылок. И если в Сети множество ссылок на сайт Microsoft, в тексте которых есть вышеперечисленные медицинские термины, то Яндекс наивно предполагает, что именно этому сайт и посвящен (похожие проблемы, кстати, и у Google — и проблемы, судя по всему, пока что неразрешимые; издержки технологии). Менее известен случай с официальным сайтом президента Беларуси, который выводился четвертым в списке результатов в ответ на запрос *жопа*»⁵². Упомянутый Ильей Google по запросу *dumb motherfucker* адресовал на страницу, посвящённую президенту Бушу. Неожиданности могли возникать и при автоматической проверке орфографии. Так, в начале века спеллингчекер Яндекса предлагал в пушкинском стихотворении заменить в слове *Пуцин* третью букву на *т*, а слово *мультиканальный* разбивал на два: *мультик анальный*.

Google, имея огромную проиндексированную базу, прекрасно находил какие-нибудь песни по 2–3 словам, но совершенно не различал омонимы, не «понимал» связь понятий *ребенок-дети* или *иду-шла-пошел*, и что при этом наречие *пошло* – совсем другое слово. Не знал он и того, что слово *простой* может быть и прилагательным, и существительным, и глаголом. Особенности обработки стоп-слов приводили к тому, что при запросе *To be or not to be* выдавалось все что угодно, кроме шекспировских текстов. А выражения *эти типы стали есть в цеху*

⁵¹ <https://www.nic.ru/whois/?searchWord=google.ru>

⁵² И.В. Сегалович. Найдется все. *Компьютерра* 29 октября 2002. <https://old.computerra.ru/200376>

или мальчик склеил модель в клубе ставили в тупик не только заокеанские, но и отечественные поисковые системы, не умевшие извлекать смысл из контекста.

Прошедшим летом и Яндекс, и Google неожиданно оказались в сходной ситуации. В июне почти синхронно два мировых суда Москвы постановили наложить на обе компании штрафы. Правда, поводы и суммы существенно различаются. Google постоянно сталкивается с обвинениями в нарушении антимонопольного законодательства из-за запрета производителям мобильных устройств предустанавливать сервисы других компаний на операционную систему Android. В этот раз в дополнение к прошлогоднему штрафу в 2 млрд рублей за «злоупотребление доминирующим положением на рынке сервисов видеохостинга YouTube» компания должна заплатить в бюджет еще более 4 миллиардов⁵³. А Яндекс оштрафован на два миллиона рублей «за непредоставление ФСБ информации о пользователях»⁵⁴. В ответ компания заявила, что установить нужное оборудование мешают санкции. Действительно, такое оборудование стоит немалых денег, и Яндексу проще и дешевле заплатить штраф, чем соблюдать требования «закона Яровой».

Надо признать, по сумме штрафов Google превосходит Яндекс с разгромным счетом. Возьмем только последние годы. Декабрь 2021 года. Штраф 7.221 млрд руб. за то, что компания не удалила информацию, признанную запрещенной. Июль 2022. 21.7 млрд руб. за «повторное неудаление недостоверной информации о военной операции на Украине»⁵⁵. Май 2023. 3 млн руб. за «неудаление роликов с пропагандой нетрадиционных сексуальных отношений, ругинга и дискредитацию ВС РФ»⁵⁶. Сегодня, 20 декабря 2023. «Назначить наказание в виде административного штрафа в размере 4 миллиардов 611 миллионов 738 тысяч 636 рублей 60 копеек»⁵⁷ (по той же статье 13.41 КоАП 23 ноября и 9 декабря – штрафы по 4 млн руб.)⁵⁸.

Начав раздел со статистики, статистикой и закончим. Приведем две

⁵³ <https://www.interfax.ru/russia/909172>

⁵⁴ <https://business.digitalocean.ru/n/yandeks-poluchil-shtraf>

⁵⁵ <https://www.rbc.ru/rbcfreenews/6582dcba9a794782f8270850>

⁵⁶ <https://ria.ru/20230511/google-1870970990.html>

⁵⁷ <https://3dnews.ru/1097760/google-snova-oshtrafovali-v-rossii-na-etot-raz-na-46-mlrd-rubley>

⁵⁸ <https://www.kommersant.ru/doc/6412720>

таблицы, характеризующие место Google и Яндекс в рунете на конец 2023 г.

	20 декабря, среда		в среднем за 7 дней	
Google	43.185.604	55.2%	47.935.747	55.0%
Яндекс	34.850.840	44.6%	39.001.663	44.8%
Rambler	46.885	0.1%	52.190	0.1%
Search.Mail.ru	38.671	0.0%	50.611	0.1%
Bing	5.248	0.0%	5.355	0.0%
Yahoo	2.183	0.0%	1.910	0.0%
Baidu	1.595	0.0%	1.890	0.0%
QIP.ru	921	0.0%	812	0.0%
Поиск.ru	471	0.0%	417	0.0%

Переходы из поисковых систем, 20.12.2023.

<https://www.liveinternet.ru/stat/ru/searches.html>

Яндекс	1 258 875 617	64.45 %
Google	684 305 035	35.03 %
Mail.ru	1 912 881	0.10 %
Rambler	688 746	0.04 %
Остальные	7 476 287	0.38 %

Поисковые системы по визитам за неделю 4—10 декабря 2023 года.

<https://radar.yandex.ru>

ДРУГИЕ

Существует множество поисковых систем, и то обстоятельство, что Google и Яндекс – бесспорно, незаменимые инструменты для поиска в сети, не означает, что другие системы не смогут найти свои нишу и аудиторию. Да, специализированные поисковые системы индексируют меньше страниц, чем гиганты, но в ряде случаев они могут точнее удовлетворить запросы пользователей. 12 апреля 2005 года была запущена Nigma (В. Лавренко и В. Чернышов) – интеллектуальная ме-

тапоисковая система, ориентированная на студентов и предназначенная для решения научных задач. Она быстро стала «широко известной в узких кругах». С помощью математического инструментария можно было решать уравнения и системы, вычислять интегралы и производные, анализировать функции и строить графики, в том числе для неявно заданных пространственных функций. Система работала с химическими терминами и формулами. Важной особенностью была возможность кластеризации результатов поиска, позволяющая фильтровать информацию. В 2012 г. Лавренко переехал во Вьетнам, где запустил поисковик Soc Soc, ставший самым популярным интернет-сайтом Вьетнама.

Среди других альтернативных поисковых систем назовем прекрасный поисковый сервис Astronet.ru – это база знаний по астрономии, созданная специалистами Государственного астрономического института им. П.К. Штернберга. Портал Medpoisk.ru не пополняется с 2016 года, но поиск по сохранившейся медицинской информации действует. Поисковая система детских сайтов Agakids.ru создана в помощь детям для поиска детских ресурсов, но полезна также и родителям (материалы по воспитанию, здоровью детей и помощи молодым мамам). Сайт Briefly.ru представляет краткое содержание многих книг. Getty Images (<https://www.gettyimages.in>) выполняет поиск по изображениям, сервис содержит также видео. Самые разнообразные звуки поможет найти FindSounds.com. Доступен интерфейс на 8 языках, в т.ч. на русском. Существуют системы, учитывающие религиозные запреты, – так, с 2013 года действует Halalgoogling – система, выдающая пользователям только халяльные «правильные» ссылки, фильтруя результаты поиска. Из других религиозно-ориентированных поисковых систем распространены Jewogle – еврейская версия Google и SeekFind.org – христианский сайт, включающий фильтры. Список можно продолжать.

В мире не так много стран, где разработаны национальные поисковые системы, не уступающие на своем рынке американскому гиганту. Это, конечно, Россия с Яндексом, не отстает и Китай. Один из крупнейших в мире поисковик Baidu основан в 2000 году Робин Ли (Robin Li Yanhong) и Эриком Сю (Eric Xu Yong). Стартовый капитал в размере \$1.2 млн предоставили венчурные компании. До этого г-н Ли стал бакалавром информатики в Пекинском университете и маги-

стром computer science в университете Нью-Йорка в Буффало. В 1996 году он придумал поисковый механизм, определявший популярность сайта на основе анализа ссылок на него. Этой разработкой заинтересовались в Infoseek, одном из пионеров поисковой индустрии. Проработав там два года, Ли со своим другом Сю вернулся в Китай, где они занялись национальным поисковиком. В 2005 году компания осуществила IPO на бирже NASDAQ, при размещении акции Baidu установили рекорд— к концу торговой сессии они подорожали в 4 с лишним раза.



Робин Ли и Эрик Сю

Помимо Baidu, в Китае есть Sogou, Shenma, Haosou/360, но они значительно уступают лидеру. Из других крупных национальных систем назовем Seznam в Чехии, Naver и Daum в Корее. Не столь успешен во Франции Qwant, его доля не превышает нескольких процентов. Отметим еще Ecosia, выпущенную в 2009 году в Германии. Особенность системы Ecosia – в ее экологической функции: 80% прибыли направляется на финансирование международного проекта по озеленению.

ЗАКЛЮЧЕНИЕ

Поисковые системы остаются среди наиболее популярных и полезных сетевых ресурсов; их разработчики и владельцы уверенно смотрят в будущее. Но по описанным объективным причинам судьба Яндекса вызывает вопросы. Отчёт [5] констатирует, что «возможность прогнозирования в краткосрочной и среднесрочной перспективе по-прежнему остаётся ограниченной из-за сохраняющейся высокой степени неопределённости в отношении геополитических и макроэкономических событий в будущем». Нам остается наблюдать за развитием ситуации и ждать новостей.

СПИСОК ЛИТЕРАТУРЫ

1. Мировая сеть Internet: применение в науке и бизнесе / Ю.М. Горностаев, Н.Н. Гурин, В.Н. Карбанов [и др.]. – Москва: Эко-Трендз, 1994. 135 с. (Технологии электронных коммуникаций; 59). EDN VLWMAN
 2. Поляк Ю.Е. Найдется все. Если уметь искать // Информационные ресурсы России. 2002. № 1-2 (64-65). С. 44–48.
 3. Поляк Ю.Е. 20 лет отечественного поиска в Интернете // Информационные технологии и математические методы в экономике и управлении (ИТИММ–2017): Сборник научных статей. Москва. 30–31 марта 2017 года. Москва: Российский экономический университет имени Г.В. Плеханова. 2017. С. 158–165. EDN WDJKNF
 4. Поляк Ю.Е., Шарова В.К. Сеть помещает людей в новые условия (Интервью с Эстер Дайсон) // Информационные ресурсы России. 2009. № 4(110). С. 2–4. EDN LAJQDZ
 5. Яндекс объявляет финансовые результаты за III квартал 2023 года [Электронный ресурс]. URL: <https://ir.yandex.ru/financial-releases?year=2023&report=q3>
-

SUBJECTIVE NOTES ON SEARCH ENGINES

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 reason for this work was the 25th anniversary of the main search engines for our country, Yandex and Google, which happened one after another. This article is an attempt to describe some events from the history of the development of Internet navigation tools from the point of view of a witness (and partly a participant).

Keywords: *Yandex, Google, search engines, search history*

REFERENCES

1. Worldwide Internet: application in science and business / Yu. Gornostaev, N. Gurin, V. N. Karbanov [and others]. Moscow: Eco-Trends, 1994. 135 p. (Electronic communications technologies; 59). EDN VLWMAN
 2. *Polak Y.E.* Everything will be found. If you know how to search // Information resources of Russia. 2002. No. 1-2 (64-65). P. 44–48. EDN TRTSKB
 3. *Polak Y.E.* 20 years of domestic search on Internet // Information technologies and mathematical methods in economics and management (ITIMM–2017): Collection of scientific articles. Moscow. March 30–31, 2017. Moscow: Russian University of Economics named after G.V. Plekhanov. 2017. S. 158–165. EDN WDJKNF
 4. *Polak Y.E., Sharova V.K.* The network places people in new conditions (Interview with Esther Dyson) // Information resources of Russia. 2009. No. 4 (110). P. 2–4. EDN LAJQDZ
 5. Yandex announces financial results for the 3rd quarter of 2023 [Electronic resource]. URL: <https://ir.yandex.ru/financial-releases?year=2023&report=q3>
-

СВЕДЕНИЯ ОБ АВТОРЕ



ПОЛЯК Юрий Евгеньевич – кандидат экономических наук, ведущий научный сотрудник Центрального экономико-математического института РАН (Москва).

Подробнее: <http://computer-museum.ru/articles/sovet-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/sovet-muzeya/561/>

email: polak@cemi.rssi.ru

ORCID: 0000-0001-8411-335X

Материал поступил в редакцию 20 декабря 2023 года

УДК 004.021, 004.42

ВИРТУАЛЬНАЯ ВЫСТАВКА КАК СРЕДСТВО ИНТЕГРАЦИИ В ЕДИНОЕ ЦИФРОВОЕ ПРОСТРАНСТВО НАУЧНЫХ ЗНАНИЙ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ В ОБЛАСТИ НАУКИ И КУЛЬТУРЫ

И. Н. Соболевская¹ [0000-0002-9461-3750], **А. Н. Сотников**² [0000-0002-0137-1255]

^{1, 2} Межведомственный суперкомпьютерный центр Российской академии наук – филиал Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»

¹ins@jsgc.ru, ²a.sotnikov@jsgc.ru

Аннотация

Рассмотрен принцип формирования виртуальных выставок как средства интеграции в Единое Цифровое Пространство Научных Знаний (ЕЦПНЗ) информационных систем в области науки и культуры с целью продвижения науки, обеспечения доступа к информации в разных областях науки, привлечения внимания к актуальным проблемам и достижениям в научной сфере. Представлены основные методы создания виртуальных выставок, включая выбор контента и разделение на основные разделы. Кроме того, предложена классификация виртуальных выставок на автономные, удаленные и комбинированные. Особое внимание уделено методологии формирования виртуальных выставок в МСЦ РАН. На примере межведомственной комбинированной виртуальной выставки предоставлено подробное описание выставки «Госпожа Пенициллин», посвященной создательнице пенициллина З. В. Ермольевой.

Ключевые слова: виртуальная выставка, Единое Цифровое Пространство Научных Знаний, Госпожа Пенициллин, связанные данные, З.В. Ермольева

ВВЕДЕНИЕ

Концепция Единого Цифрового Пространства Научных Знаний предполагает создание централизованной информационной среды, которая обеспечит доступ к информации в различных областях науки. Цели и задачи этой концепции, вклю-

чая популяризацию научных достижений, указаны в [1–5]. Цели и задачи создания ЕЦПНЗ, перечисленные в [1], включают, наряду с чисто научными и образовательными, популяризацию научных достижений.

Одним из эффективных инструментов для популяризации науки являются виртуальные выставки (ВВ). Такие выставки предоставляют широкой аудитории пользователей интернета возможность познакомиться с историей науки, новейшими достижениями в конкретных научных областях, а также проверить свои знания в интерактивной форме.

Виртуальные выставки являются эффективным инструментом для популяризации науки, предоставляя возможность интерактивного ознакомления с научными достижениями и историей науки. Многие музеи мира и крупные национальные библиотеки предлагают посетителям своих сайтов виртуальные туры или цифровые коллекции своих фондов (например, <https://naturalhistory.si.edu/visit/virtual-tour>, <https://hsm.ox.ac.uk/collections>, <https://www.louvre.fr/visites-en-ligne?defaultView=rdc.s46.p01&lang=ENG>, <https://www.loc.gov/collections/>). Пользователям предлагается, например, познакомиться с некоторыми экспонатами музея, изучить тематические подборки архивных документов и печатных изданий или исследовать историю здания. Однако такие выставки ограничены материалами одной организации.

Формирование межведомственных виртуальных выставок и их интеграция в Единое Цифровое Пространство Научных Знаний позволят предоставить пользователю более разнообразную информацию по данной тематике. ВВ, сформированные путем объединения разнородной информации и собранные на единой платформе из различных институтов памяти, способствуют созданию более широкой доступности знаний для пользователей. Интеграция в ЕЦПНЗ виртуальных выставок, охватывающих разнообразные тематики и музейные объекты, способствует более полному и глубокому пониманию научных и культурных явлений.

Поскольку виртуальные выставки становятся все более популярными, необходимо рассмотреть основные принципы их создания.

1. ПРИНЦИПЫ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ВЫСТАВОК

ВВ создаются с использованием современных технологий визуализации, интерактивности и 3D-моделирования. 3D-моделирование позволяет более полно и детально представить объекты, которые могут быть изучены с разных ракурсов, что повышает их наглядность и удобство для изучения. Кроме того, технологии визуализации и интерактивности значительно повышают уровень эффекта «присутствия» на выставке (как, например, в проекте, посвященном научному наследию М.М. Герасимова, создателя метода пластической портретной реконструкции (<http://exhibitions.jssc.ru/gerasimov/index.php/dokumenty-epokhi>)).

ВВ могут быть как временными объединениями разнородных ресурсов (по аналогии с временными выставками в музеях), так и ресурсами постоянного хранения (как постоянная экспозиция). Постоянные ВВ, если они не носят сугубо исторического характера, могут дополняться новой информацией по мере ее появления.

Формирование виртуальных выставок — это процесс создания онлайн-платформы, на которой можно просмотреть и изучить разнородные цифровые ресурсы, относящиеся к определенному событию, персоне, научному направлению. Виртуальные выставки создаются на основе фотографий, 3D моделей объектов, текстовых, аудио- и видеоматериалов. Кроме ресурсов этих видов, они могут включать и интерактивную компоненту.

Интерактивная компонента ВВ вовлекает пользователя в диалоговый процесс исследования и ознакомления с тематикой выставки. Она позволяет пользователю участвовать в игровых элементах и тестировании, проверяя свои знания и понимание тематики выставки; общаться с другими пользователями, делаясь своими впечатлениями и знаниями.

2. ВЫБОР КОНТЕНТА ВИРТУАЛЬНОЙ ВЫСТАВКИ

Выбор контента для виртуальной выставки зависит от ее тематики. Однако существует несколько общих принципов, которые могут быть при этом применены:

1. Доступность. Контент выставки должен представлять интерес для широкой аудитории и состоит, как правило, из разнородной информации (текстовые документы, аудио-, видеофайлы и т. д.).

2. Контент может предоставляться различными организациями и/или физическими лицами.

3. Контент не должен содержать составляющих, не соответствующих тематике ВВ.

4. Актуальность. Контент ВВ не должен содержать устаревшую информацию.

5. Разнообразие. Контент виртуальной выставки не должен быть однородным. Рекомендуется включение разнообразных видов контентных объектов, таких как видео, статические изображения и интерактивные элементы.

3 СТРУКТУРА ВИРТУАЛЬНОЙ ВЫСТАВКИ

Ресурсы, включаемые в ВВ, представляют собой контент с различными форматами представления данных (jpeg, avi, pdf, txt, mp3 и т. д.). Кроме того, существуют разные стандарты описания тех или иных видов объектов: например, для описания и классификации экспонатов могут использоваться Dublin Core, MIDAS, MARC и др. Поэтому структура ВВ должна иметь архитектуру, позволяющую обрабатывать и визуализировать различные типы данных и форматы, а также обеспечивать высокую производительность и безопасность работы. Структура ВВ должна иметь следующие основные свойства:

1. Расширяемость. Структура выставки должна позволять добавлять новые тематические разделы, видео- и аудио контент, фотографии, игровые элементы и т. д.

2. Гибкость. Структура ВВ должна позволять при необходимости вносить изменения в размещение информации, адаптировать сайт под новые возможности и функции.

4 ОСНОВНЫЕ РАЗДЕЛЫ ВИРТУАЛЬНОЙ ВЫСТАВКИ

ВВ может быть посвящена, как уже указывалось, различным видам объектов и включать широкий круг специфических разделов. Можно выделить несколько типовых разделов ее контента:

- Биография персоны/персон (если ВВ посвящена ученому/ученым) / история формирования научного направления;
- Библиотека, включающая библиографию и цифровые копии печатных изданий, относящихся к теме ВВ;
- Архивные документы;
- Мультимедийные материалы (фото-, кино-видео-, аудиоматериалы, 3D-модели);
- Научно-популярный раздел, содержащий познавательную информацию по теме выставки, рассчитанную на широкий круг пользователей. Этот раздел может также включать интерактивную составляющую, например, онлайн викторину или игру, направленную на изучение материалов по теме ВВ;
- Информация о выставке, ее авторах, спонсорах и т. п.;
- Контакты, форма обратной связи с организаторами ВВ.

5 ТИПЫ ВИРТУАЛЬНЫХ ВЫСТАВОК

ВВ включают не только информацию, физически хранящуюся на компьютерах, поддерживающих выставки, но и ссылки на внешние ресурсы. ВВ, также как и ЕЦПНЗ, включают элементы, обеспечивающие:

- - доступ к внешним ресурсам;
- - визуализацию данных;
- - поиск и фильтрацию информации;

и т. д. [6].

ВВ так же, как и ЕЦПНЗ в целом, содержат разнородную информацию, собранную из разных источников. При этом такая информация может:

- храниться на одном сервере; в этом случае будем называть ВВ автономной;
- предоставляться онлайн в процессе работы пользователя с ВВ и храниться на серверах держателей ресурсов (библиотек, музеев, архивов и т. п.); в этом случае будем называть ВВ удаленной;
- быть «комбинированной»: часть информации хранится на сервере ЕЦПНЗ, а часть – на серверах держателей информации; в этом случае будем называть ВВ комбинированной.

6 МЕТОДОЛОГИЯ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ВЫСТАВОК В МСЦ РАН

В основу технологии формирования ВВ положен принцип распределенного формирования и централизованной поддержки ВВ, аналогичный оправдавшим себя принципам формирования ЭБ «Научное наследие России» [7, 8].

Основные принципы формирования ВВ рассмотрим на примере ВВ «Госпожа Пенициллин», посвященной З.В. Ермольевой и ее научной школе (<http://exhibitions.jssc.ru/ermoleva/>) (Рис. 1).



Рис. 1. Главная страница ВВ «Госпожа Пенициллин».

Эта ВВ представляет собой ресурс комбинированного типа. Она объединяет в одном пространстве информацию, разбросанную по разным организациям и частным лицам (выставка создана в рамках совместного проекта 12 организаций, включая библиотеки, музеи, архивы, научно-образовательные организации. список которых представлен в одном из разделов выставки (см. ниже)), многоаспектную разнородную информацию, относящуюся к жизни и деятельности Е.В. Ермольевой, и предоставляет пользователям возможности ознакомления с ней через единую точку входа. Разделы выставки содержат текстовые материалы, оцифрованные книги, фотографии и архивные документы, цифровые копии полнометражных художественных фильмов, фрагменты научно-популярных и документальных фильмов, связанных с тематикой выставки. На этой выставке также представлены ссылки на внешние источники, содержащие видео контент и текстовые публикации.

7 «ГОСПОЖА ПЕНИЦИЛЛИН»

Выставка содержит 6 основных разделов:

- Биография;
- Мультимедиа;
- Библиотека;
- Это интересно;
- О проекте;
- Ермольева в ННР.

Каждый раздел, кроме раздела «Ермольева в ННР», содержит подразделы.

Раздел «Биография»

Кроме общей биографической справки, в этом разделе представлена информация об основных направлениях научных интересов З.В. Ермольевой (Рис. 2), ее преподавательской деятельности, а также информация о памятных объектах, связанных с именем Ермольевой (Рис. 3).

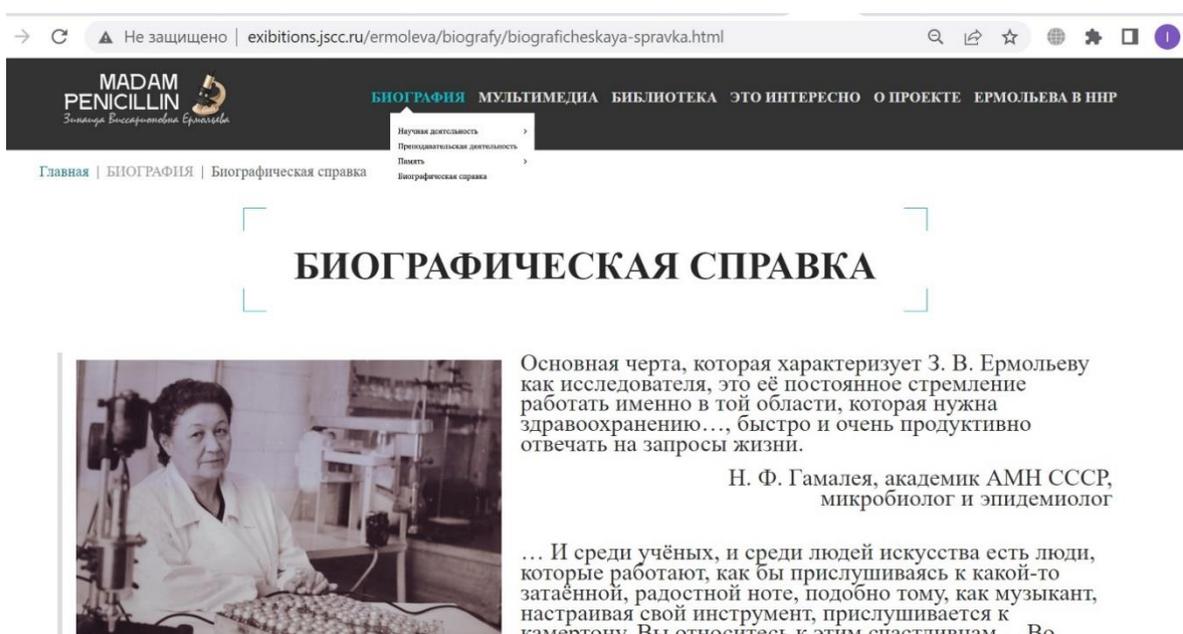


Рис. 2. Основные направления научных исследований З.В. Ермольевой. Биографическая справка.

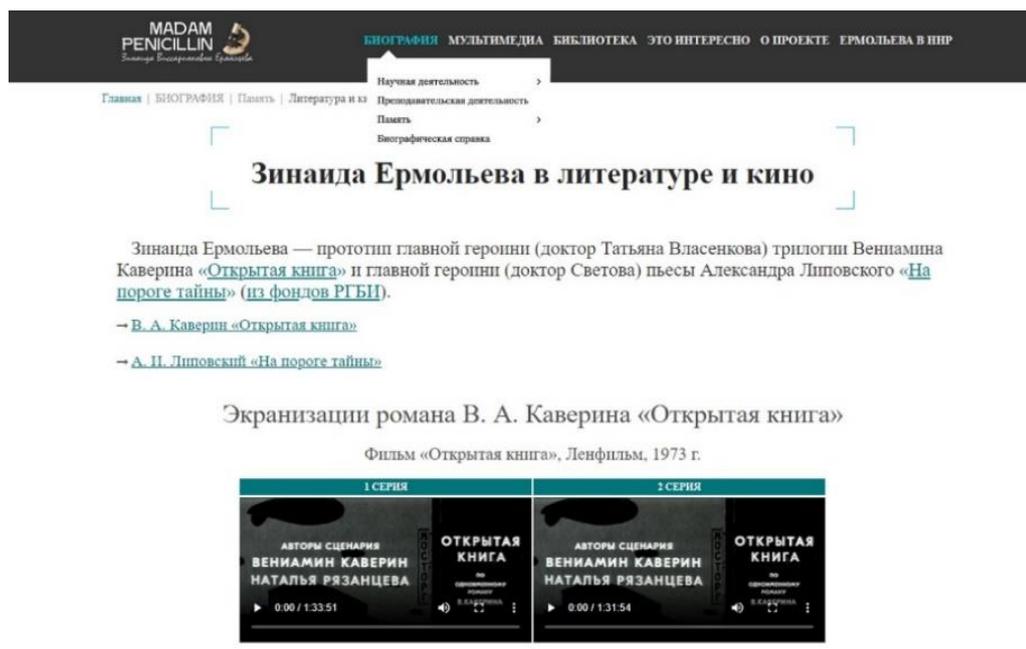


Рис. 3. Память о З.В. Ермолевой.

Раздел «Мультимедиа» содержит 3 подраздела: фотоархив, видеоархив и архивные документы, связанные с З.В. Ермолевой (Рис. 4).

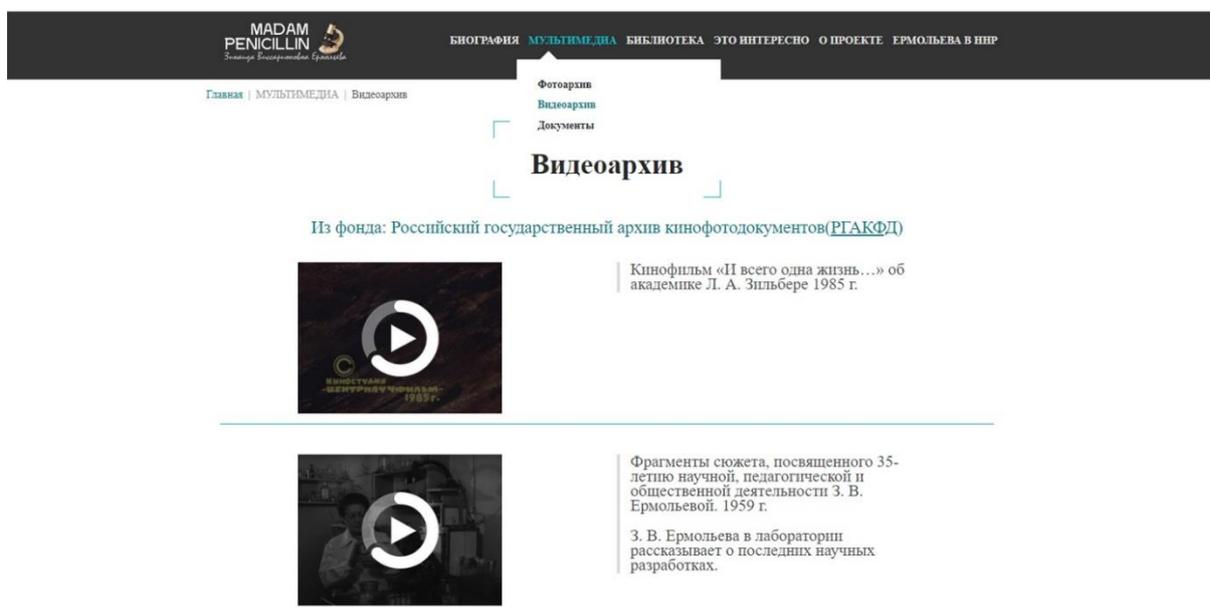


Рис. 4. Содержание раздела «Мультимедиа».

В этом разделе содержатся, в частности, цифровая копия рукописной автобиографии З.В. Ермолевой, написанной ею в 1974 году, письмо врача эвакогоспиталя

Потемкина З.В. Ермольевой с благодарностью за спасение жизни раненого с помощью пенициллина. г. Москва, 1943 г.

Раздел «Библиотека» содержит 3 подраздела: публикации З.В. Ермольевой, публикации о З.В. Ермольевой, Микробиология (Избранное). В этом разделе содержатся библиография работ о З.В. Ермольевой со ссылками на ряд публикаций, представленных полными текстами на внешних источниках, публикации, посвященные микробиологии, труды З.В. Ермольевой (Рис. 5). Цифровые копии печатных изданий представлены в формате pdf, с возможностью масштабирования изображения.

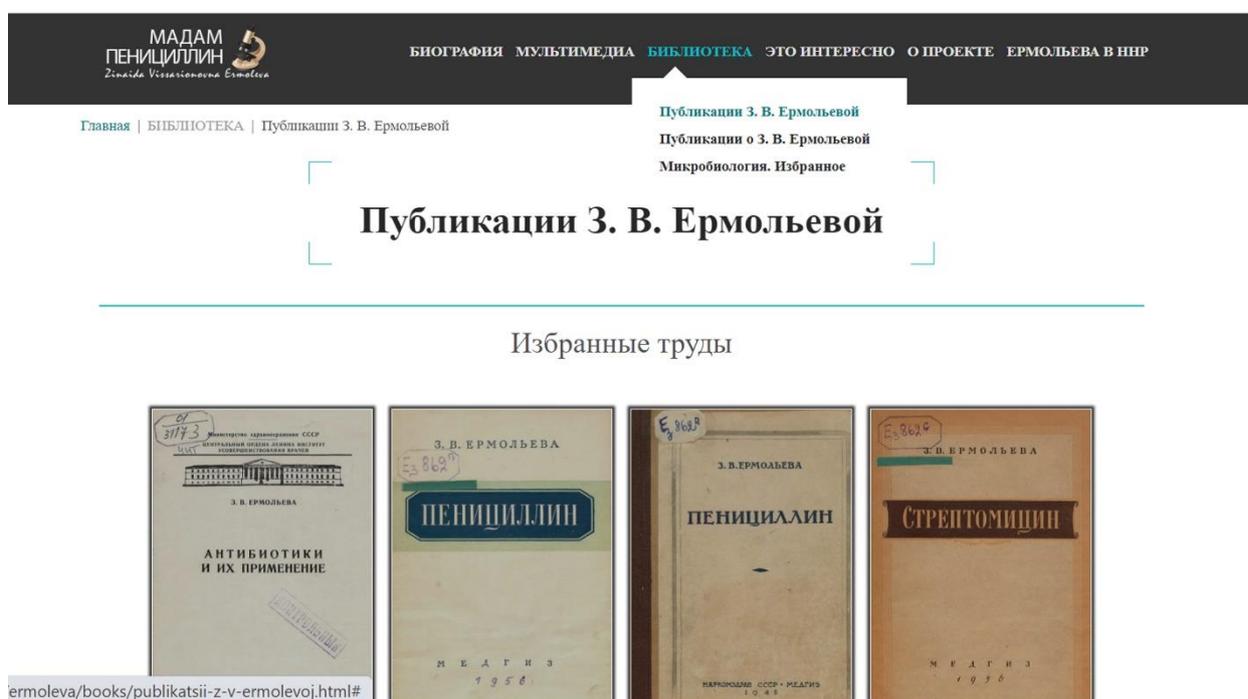


Рис. 5. Содержание раздела «Библиотека».

Раздел «Это интересно»

Этот раздел содержит научно-популярную информацию о микробиологии, викторину по микробиологии, позволяющую в интерактивной форме познакомиться с некоторой информацией по микробиологии, а также раздел «География Ермольевой» (Рис. 6).



Рис. 6. Содержание раздела «Это интересно».

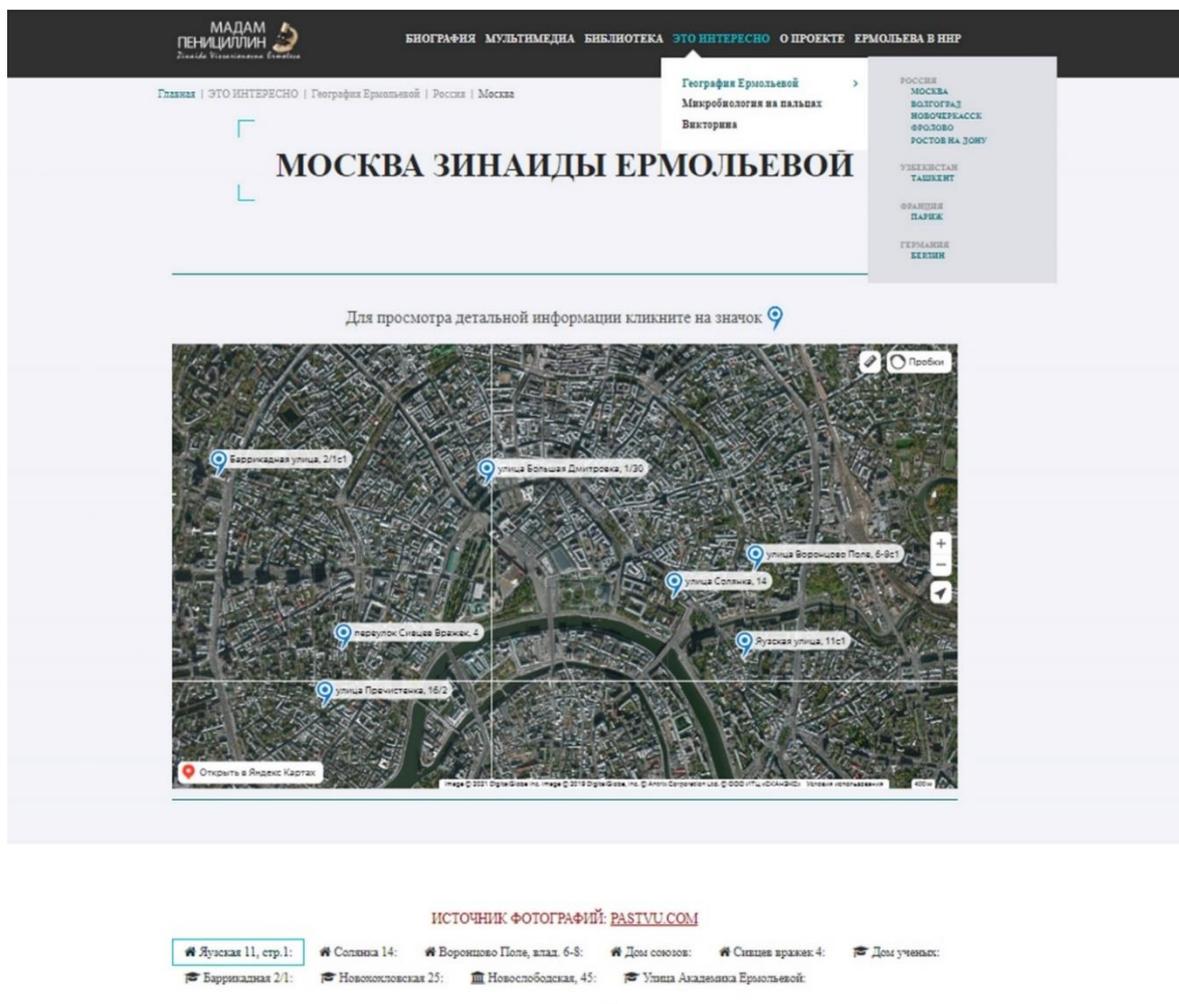


Рис. 7. Содержание раздела «География Ермольевой».

«География Ермольевой»

З.В. Ермольева, в силу своей научной деятельности, много путешествовала. В этом разделе представлены адреса на картах городов, так или иначе связанные с Ермольевой (Рис. 7). В частности, «Москва Ермольевой» представлена в виде фотографий старой Москвы, взятых с общедоступного сайта <https://pastvu.com/>, а также интерактивной картой, содержащей информацию о существующих (или существовавших ранее) архитектурных объектах, в той или иной мере связанных с З.В. Ермольевой.

Раздел «О проекте»

В этом разделе представлены участники проекта и контакты для обратной связи (Рис. 8).

MADAM
PENICILLIN
Зинаида Виссарионовна Ермольева

БИОГРАФИЯ МУЛЬТИМЕДИЯ БИБЛИОТЕКА ЭТО ИНТЕРЕСНО **О ПРОЕКТЕ** ЕРМОЛЬЕВА В ННР

Главная | О ПРОЕКТЕ | Участники проекта

О Проекте

Виртуальная выставка, посвящённая жизни и научному наследию Зинаиды Виссарионовны Ермольевой, собирает в одном пространстве разбросанную в настоящий момент по разным организациям и частным лицам информацию и делает её доступной для любого заинтересованного читателя и зрителя. Этот проект — пример интеграции информационных ресурсов библиотек, научных учреждений, архивов и музеев. Он открыт для всех потенциальных партнёров, число которых, как мы надеемся, будет увеличиваться.

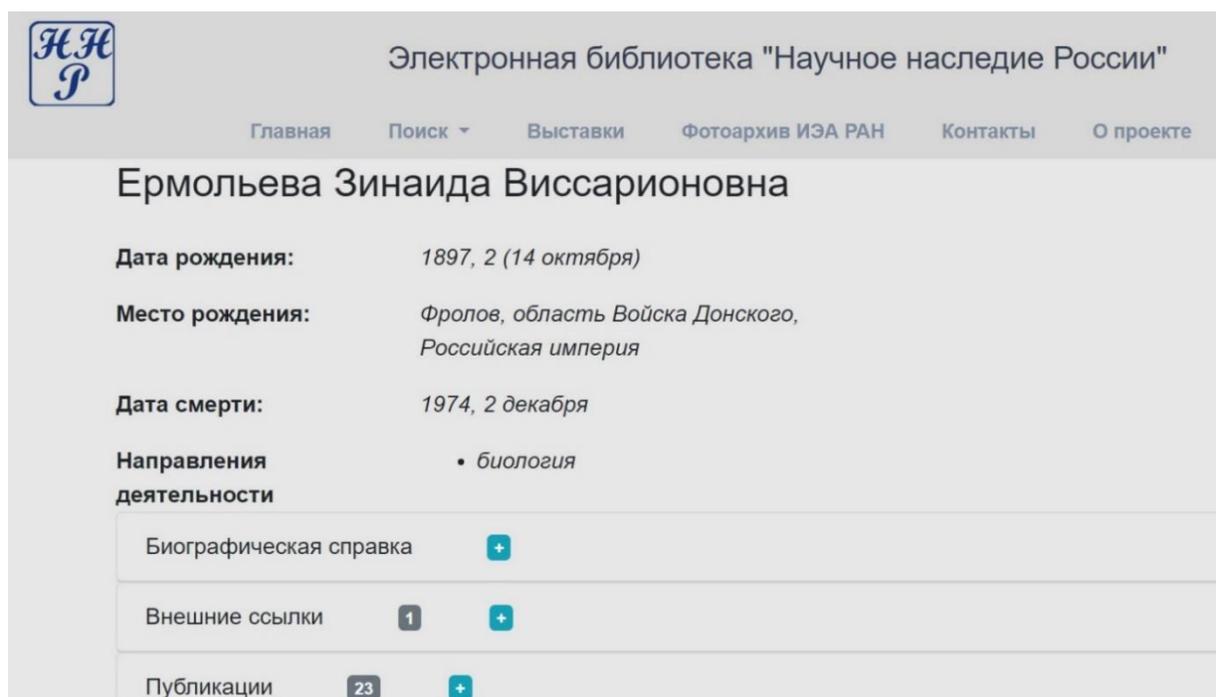
Участники проекта

- [Российская государственная библиотека искусств \(РГБИ\)](#)
- [Военно-медицинский музей Министерства обороны Российской Федерации \(ВММ\)](#)
- [Государственный биологический музей им. К. А. Тимирязева \(ГБМТ\)](#)
- [Государственный центральный музей современной истории России \(ГЦМСИР\)](#)
- [Межведомственный суперкомпьютерный центр РАН](#)
- [Национальная Академия Микологии](#)
- [Российский государственный архив кинофотодокументов \(РГАКФД\)](#)
- [Российский государственный архив экономики \(РГАЭ\)](#)
- [Российская медицинская академия непрерывного профессионального образования \(РМАНПО\)](#)
- [ФИЦ Биотехнологии РАН](#)
- [Фроловский Городской краеведческий музей \(Волгоградская область\)](#)
- [Центральная научная медицинская библиотека](#)

Рис. 8. Содержание раздела «О проекте».

Раздел «Ермольева в ЭБ ННР»

Переходя в этот раздел, пользователь попадает на страницу З.В. Ермольевой в ЭБ ННР (Рис. 9). В ЭБ ННР также представлены биографическая справка о З.В. Ермольевой, ссылки на внешние источники, связанные с персоной, цифровые копии некоторых ее трудов.



Электронная библиотека "Научное наследие России"

Главная Поиск Выставки Фотоархив ИЭА РАН Контакты О проекте

Ермольева Зинаида Виссарионовна

Дата рождения: 1897, 2 (14 октября)

Место рождения: Фролов, область Войска Донского, Российская империя

Дата смерти: 1974, 2 декабря

Направления деятельности • биология

- Биографическая справка +
- Внешние ссылки 1 +
- Публикации 23 +

Рис. 9. З.В. Ермольева в ЭБ ННР.

За период с 01.03.2023 по 01.09.2023 сайт виртуальной выставки «Госпожа Пенициллин» в среднем посещали 117 чел./день.

ЗАКЛЮЧЕНИЕ

Каждый музей обладает набором отличительных признаков, которыми определяется его направленность. В каждом музее есть свои методические рекомендации по описанию объекта.

В настоящее время ни один музей не может обойтись без использования информационных систем и баз данных, специально разработанных для нужд культурных учреждений. Эти системы предназначены для обеспечения сохранности оригиналов экспонатов, улучшения доступа к информации о коллекциях му-

зая, быстрого поиска необходимых данных, удобного учета экспонатов и получения прибыли от цифровой продукции [9].

Основу системы хранения данных музейных информационных систем составляют:

- устройства хранения;
- инфраструктура доступа к устройствам хранения;
- подсистема резервного копирования и архивирования данных;
- программное обеспечение управления хранением;
- система управления и контроля.

Интеграция системы хранения данных музейных информационных систем в единый информационный ресурс позволит эффективно управлять и обмениваться информацией. Такая интеграция способствует повышению общей эффективности управления как музейными, так и библиотечными и архивными данными и ресурсами.

Проект ВВ «Госпожа Пенициллин» — пример интеграции информационных ресурсов библиотек, научных учреждений, архивов и музеев. Он открыт для всех потенциальных партнёров, число которых, как мы надеемся, будет увеличиваться. Эта ВВ включает в себя централизованную информационную систему, объединяющую ресурсы различных учреждений. Выставка предоставляет доступ к уникальным материалам по микробиологии, а также к научному наследию З.В. Ермольевой.

Данная ВВ является примером успешного интегрирования разнородных информационных ресурсов в единый сервис, что позволяет пользователям получить доступ к уникальным материалам и использовать их в своей работе.

Накопленный опыт формирования и использования ВВ в МСЦ РАН учитывается при проектировании ЕЦПНЗ и, наоборот, – при создании новых ВВ, в проектировании которых участвует МСЦ РАН, используются принципы объединения разнородных научных ресурсов, принятые в ЕЦПНЗ, что позволяет рассматривать создаваемые ВВ в качестве моделей элементов ЕЦПНЗ, на которых отрабатываются интерфейсы взаимодействия с пользователями, методы визуализации и навигации по связанным ресурсам.

Работа выполнена в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН в рамках государственного задания № FNEF-2024-0014.

СПИСОК ЛИТЕРАТУРЫ

1. *Савин Г.И.* Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России. 2020. № 5. С. 3–5.
<https://doi.org/10.51218/0204-3653-2020-5-3-5>
2. *Антопольский А.Б., Басов А.В., Савин Г.И., Сотников А.Н., Цветкова В.А., Каленов Н.Е., Серебряков В.А., Ефременко Д.В.* Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.
<https://doi.org/10.36535/0548-0019-2020-04-2>
3. *Каленов Н.Е., Сотников А.Н.* Архитектура единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 5–8. <https://doi.org/10.51218/0204-3653-2020-5-5-8>
4. *Атаева О.М., Каленов Н.Е., Серебряков В.А.* Онтологический подход к описанию единого цифрового пространства научных знаний // Электронные библиотеки. 2021. Т. 24, № 1. С. 3–19.
<https://doi.org/10.26907/1562-5419-2021-24-1-3-19>
5. *Антопольский А.Б., Басов А.В., Савин Г.И., Сотников А.Н., Цветкова В.А., Каленов Н.Е., Серебряков В.А., Ефременко Д.В.* Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17
6. *Кириллов С.А., Соболевская И.Н., Сотников А.Н.* Принципы формирования и представления междисциплинарных коллекций в цифровом пространстве научных знаний // Электронные библиотеки. 2021. Т. 24, № 2. С. 294–314.
7. *Каленов Н.Е., Савин Г.И., Серебряков В.А., Сотников А.Н.* Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы. 2012. Т. 4, № 100. С. 30–40.
8. *Каленов Н.Е., Погорелко К.П., Сотников А.Н.* О развитии электронной библиотеки «Научное наследие России» как составляющей Единого цифрового

пространства научных знаний // Информационные процессы. 2022. Т. 22, № 3. С. 155–166. https://doi.org/10.53921/18195822_2022_22_3_155

9. Гришина Н.В. Музейные информационные системы и информационные риски их использования // Вестник МФЮА. 2018. №4. С. 96–101.

VIRTUAL EXHIBITION AS A MEANS OF INTEGRATING INTO A UNIFIED DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE AND INFORMATION SYSTEMS IN THE FIELD OF SCIENCE AND CULTURE

I. N. Sobolevskaya¹ [0000-0002-9461-3750], **A. N. Sotnikov**² [0000-0002-0137-1255]

^{1,2} *Joint Supercomputer Center of the Russian Academy of Sciences — Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences” (JSCC RAS — Branch of SRISA), 119334, Moscow, Leninsky av., 32 a, Russia*

¹*ins@jssc.ru*, ²*ASotnikov@jssc.ru*

Abstract

The study examines the principle of creating virtual exhibitions as a means of integration into the Common Digital Space of Scientific Knowledge (CDSSK), information systems in the field of science and culture, with the aim of promoting science, ensuring access to information in various scientific fields, and drawing attention to current issues and achievements in the scientific sphere. The main methods of creating virtual exhibitions are formulated, including content selection and segmentation into main sections. In addition, a classification of virtual exhibitions into autonomous, remote, and combined is proposed. Special attention is paid to the methodology of creating virtual exhibition at the Moscow Center of the Russian Academy of Sciences. Using the example of an interdepartmental combined virtual exhibition, a detailed description of the "Mrs. Penicillin" exhibition dedicated to the creator of penicillin, Z.V. Ermolyeva, is provided.

Keywords: *virtual Exhibition, Common Digital Space of Scientific Knowledge, Madame. Penicillin, related data, Z.V. Yermolyeva*

REFERENCES

1. Savin G.I. Edinoe cifrovoe prostranstvo nauchnyh znaniy: celi i zadachi // Informacionnye resursy Rossii. 2020. № 5. S. 3–5.
<https://doi.org/10.51218/0204-3653-2020-5-3-5>
2. Antopol'skij A.B., Basov A.V., Savin G.I., Sotnikov A.N., Cvetkova V.A., Kalenov N.E., Serebryakov V.A., Efremenko D.V. Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy (ECPNZ) // Nauchno-tehnicheskaya informaciya. Ser. 1. 2020. № 4. S. 9–17.
<https://doi.org/10.36535/0548-0019-2020-04-2>
3. Kalenov N.E., Sotnikov A.N. Arhitektura edinogo cifrovogo prostranstva nauchnyh znaniy // Informacionnye resursy Rossii. 2020. № 5. S. 5–8.
<https://doi.org/10.51218/0204-3653-2020-5-5-8>
4. Ataeva O.M., Kalenov N.E., Serebryakov V.A. Ontologicheskij podhod k opisaniyu edinogo cifrovogo prostranstva nauchnyh znaniy // Russian Digital Library Journal. 2021. T. 24, № 1. S. 3–19.
<https://doi.org/10.26907/1562-5419-2021-24-1-3-19>
5. Antopol'skij A.B., Bosov A.V., Savin, G.I. Sotnikov A.N., Cvetkova V.A., Kalenov N.E., Serebryakov V.A., Efremenko D.V. Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy (ECPNZ) // Nauchno-tehnicheskaya informaciya. Ser. 1. 2020. № 4. S. 9–17.
6. Kirillov S.A., Sobolevskaya I.N., Sotnikov A.N. Principy formirovaniya i predstavleniya mezhdisciplinarnykh kollekcij v cifrovom prostranstve nauchnyh znaniy // Russian Digital Library Journal. 2021. T. 24, № 2. S. 294–314.
7. Kalenov N.E., Savin G.I., Serebryakov V.A., Sotnikov A.N. Principy postroeniya i formirovaniya elektronnoj biblioteki "Nauchnoe nasledie Rossii" // Programmnye produkty i sistemy. 2012. T. 4, № 100. S. 30–40.
8. Kalenov N.E., Pogorelko K.P., Sotnikov A.N. O razvitii elektronnoj biblioteki "Nauchnoe nasledie Rossii" kak sostavlyayushchej Edinogo cifrovogo prostranstva nauchnyh znaniy // Informacionnye processy. 2022. T. 22, № 3. S. 155–166.
https://doi.org/10.53921/18195822_2022_22_3_155
9. Grishina N.V. Muzejnye informacionnye sistemy i informacionnye riski ih ispol'zovaniya // Vestnik MFYUA. 2018. №4. S. 96–101.

СВЕДЕНИЯ ОБ АВТОРАХ



СОБОЛЕВСКАЯ Ирина Николаевна – старший научный сотрудник Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», кандидат физико-математических наук.

Irina Nikolaevna SOBOLEVSKAYA – higher senior officer of Joint Super Computer Center of the Russian Academy of Sciences – Branch of Federal State Institution «Scientific Research Institute for System Analysis of the Russian Academy of Sciences», Candidate of Physics and Math Sciences.

email: ins@jscs.ru;

ORCID: 0000-0002-9461-3750



СОТНИКОВ Александр Николаевич – заместитель директора Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», доктор физико-математических наук, профессор.

Alexander Nikolaevich SOTNIKOV – Deputy Director of the Joint SuperComputer Center of the Russian Academy of Sciences – Branch of the Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, Doctor of Sciences (Math), Professor.

email: asotnikov@jscs.ru;

ORCID: 0000-0002-0137-1255

Материал поступил в редакцию 10 декабря 2023 года

УДК 519.6, 519.2

СТОХАСТИЧЕСКИЙ АНАЛИЗ ПОТОКОВ ТЕПЛА «ОКЕАН –АТМОСФЕРА» В СЕВЕРНОЙ АТЛАНТИКЕ

Н. П. Тучкова¹ [0000-0001-5357-9640], К. П. Беляев² [0000-0003-2111-2709],

Г. М. Михайлов³ [0000-0002-4535-7180]

¹⁻³Вычислительный центр им. А.А. Дородницына ФИЦ Информатика
и управление РАН, г. Москва

²Институт океанологии им. П.П. Ширшова РАН, г. Москва

¹natalia_tuchkova@mail.ru, ²kosbel55@gmail.com, ³gmickail@ccas.ru,

Аннотация

Проанализированы данные наблюдений в Северной Атлантике, полученные за 40 лет в рамках проекта NAAD. Рассматривался полный поток тепла из океана в атмосферу (и из атмосферы в океан) как сумма скрытого и явного тепла. Коэффициенты стохастического дифференциального уравнения, представляющего стохастический процесс, были статистически определены из исходного набора данных. Ранее существование и единственность решения в сильном смысле стохастического дифференциального уравнения, порожденного построенным диффузионным процессом, были доказаны при выполнении условий Колмогорова. Численные расчеты проведены на суперкомпьютере «Ломоносов-2» МГУ им. М.В. Ломоносова.

Ключевые слова: анализ временных рядов, климатический сезонный ход, максимальные и минимальные значения тепловых потоков внутри климатического года

ВВЕДЕНИЕ

В настоящей работе представлены результаты моделирования пространственно-временного распределения тепловых потоков в Северной Атлантике [1] и проведен их вероятностный анализ. Данные потоков взяты из открытого для научных использований варианта Атласа международного проекта NAAD, размещенного по адресу <https://noaadata.apps.nsidc.org/> [2]. Область моделирования (<https://naad.ocean.ru/>) проиллюстрирована на Рис. 1.

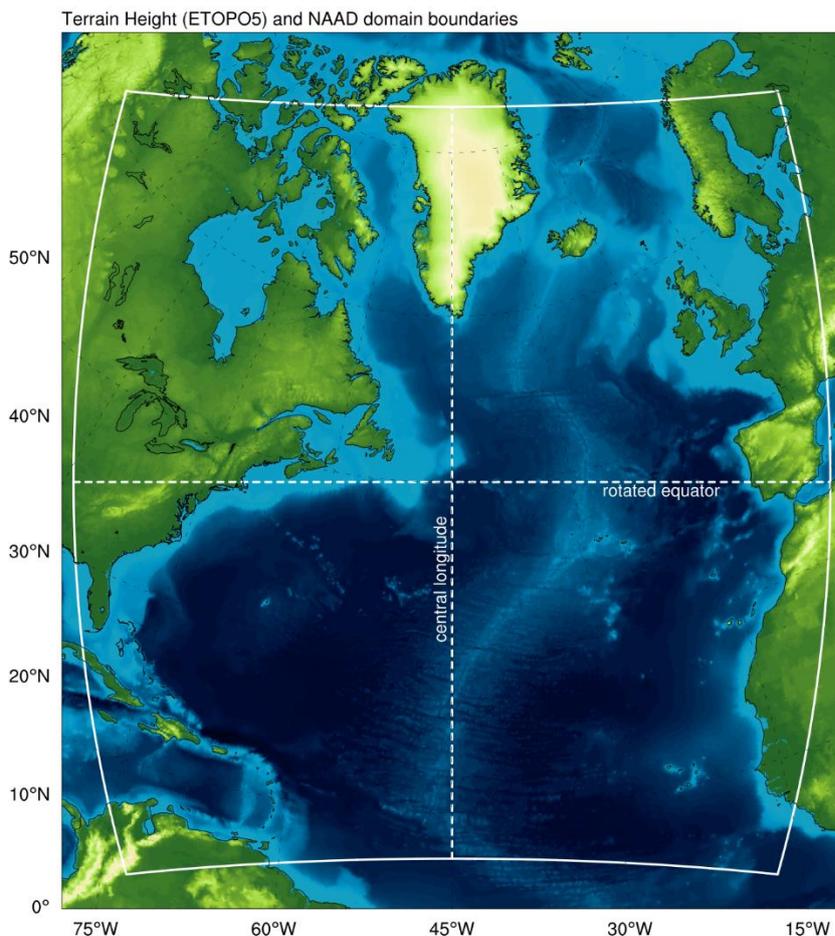


Рис. 1. Зона прогноза атмосферы Северной Атлантики (10°–80° с.ш.) за 40 лет проекта NAAD(<https://naad.ocean.ru/datarequest.php>).

Изменение потоков моделировалось с помощью стохастического дифференциального уравнения (уравнение Ито–Ланжевена)

$$dX = a(t, X)dt + b(t, X)dW, \quad (1)$$

где X – поток тепла в точке плоскости x в момент времени t , а dW – стандартное обозначение «Гауссовского белого шума». Рассматриваем более общий случай по отношению к стандартному уравнению Ланжевена [3], когда коэффициент $b(t, x)$ – случайная величина и зависит от потока.

Статистическими методами [4] определялись коэффициенты уравнения (1): $a(t, x)$ – коэффициент сноса, $b(t, x)$ – коэффициент диффузии. Строились временные и пространственные распределения этих величин. На рисунках 2 и 3 приведены графики временного хода за 1979, 1999 и 2018 годы для средних значений коэф-

коэффициентов $a(t, x)$ и $b(t, x)$. Ниже представлены краткий анализ полученных коэффициентов и их физический смысл.

Рисунки 2 (а–в) показывают поведение коэффициента сноса $a(t, x)$ в течение времени (t изменяется через сутки в течение года, а значение x определяется как среднее значение потока по рассматриваемой области в заданный момент времени, то есть среднее по области за фиксированные сутки). На рисунках 2(а–в) приведены соответствующие значения за 1979, 1999 и 2018 годы. Из представленных иллюстраций видно, что сглаженное значение этих коэффициентов хорошо аппроксимируется тригонометрической зависимостью с годовой гармоникой, как на Рис. 2 (а), и несколькими гармониками, как на Рис. 2 (б) и 2 (в). При этом на Рис. 2 (а) амплитуда основной гармоники положительна, а на остальных рисунках отрицательна. Это позволяет смоделировать климатическую (долговременную) изменчивость соответствующих коэффициентов как

$$a(t, x) = A(x) \exp(it\omega), \quad (2)$$

где $A(x)$ – амплитуда основной (как правило, годовой) гармоники, зависящей только от значения потока; ω – соответственно, частота основной гармоники, $i = \sqrt{-1}$.

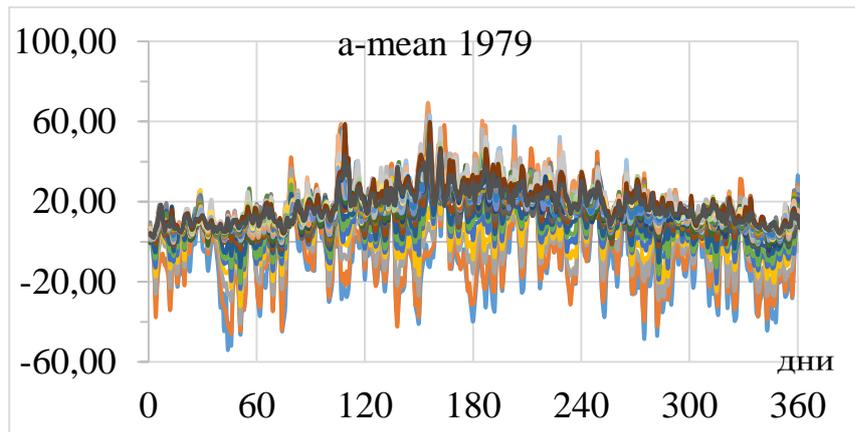
В уравнении (2) можно рассматривать более общую модель вида

$$a(t, x) = \sum_{j=1}^N A_j(x) \exp(i\omega_j t),$$

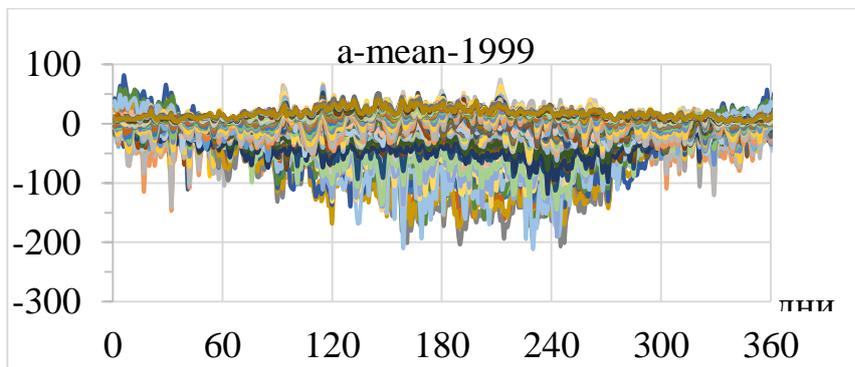
где коэффициенты $A_j(x)$ также подбираются методом наименьших квадратов, как это сделано в работе [5].

Аналогичный анализ можно сделать относительно коэффициента диффузии $b(t, x)$, показанного на Рис. 3 (а, б, в). На Рис. 3 (а) показано поведение коэффициента $b(t, x)$ за 1979 год, а на Рис. 3 (б) и 3 (в) – за 1999 и 2018 годы, соответственно. Так же, как для Рис. 2, коэффициент $b(t, x)$ рассчитывался посуточно при значении потока, равном среднему по области.

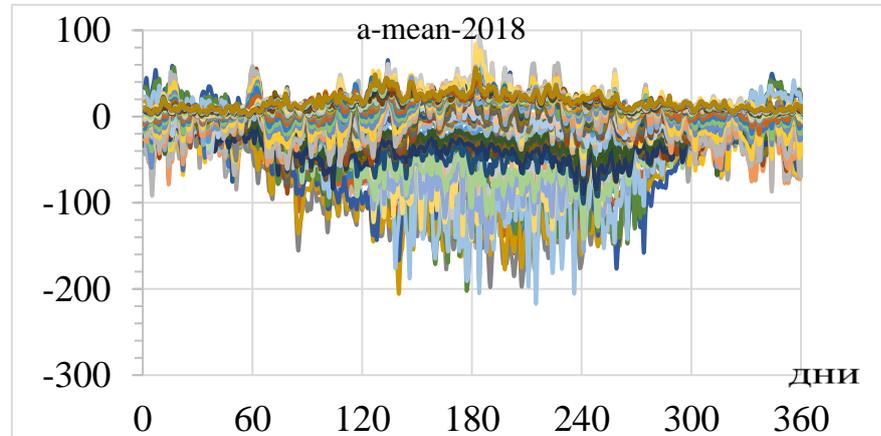
Из Рис. 3 хорошо видно, что при сглаживании коэффициент $b(t, x)$ также хорошо моделируется формулой $b(t, x) = B(x) \exp(it\omega)$, при этом значения коэффициента всегда положительны.



(а)



(б)



(в)

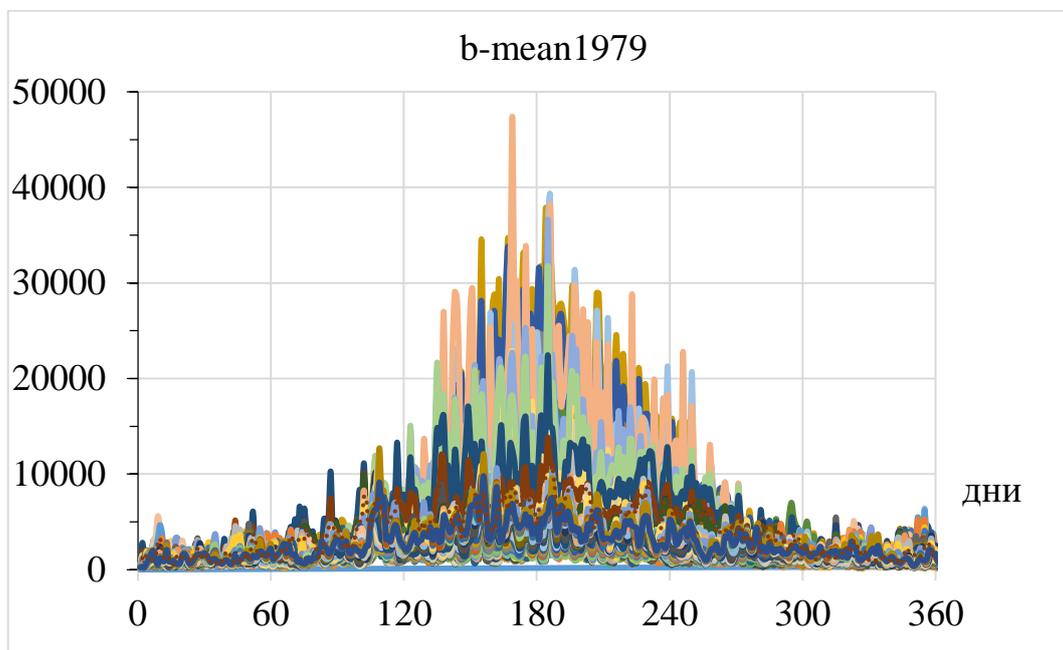
Рис. 2. Изменение средних значений коэффициента $\alpha(t,x)$ с шагом 1 день за
а) 1979; б) 1999; в) 2018 годы

Как следствие, можно написать уравнение для расчета вероятностей значе-
ний потока в заданный (климатический) момент времени в виде

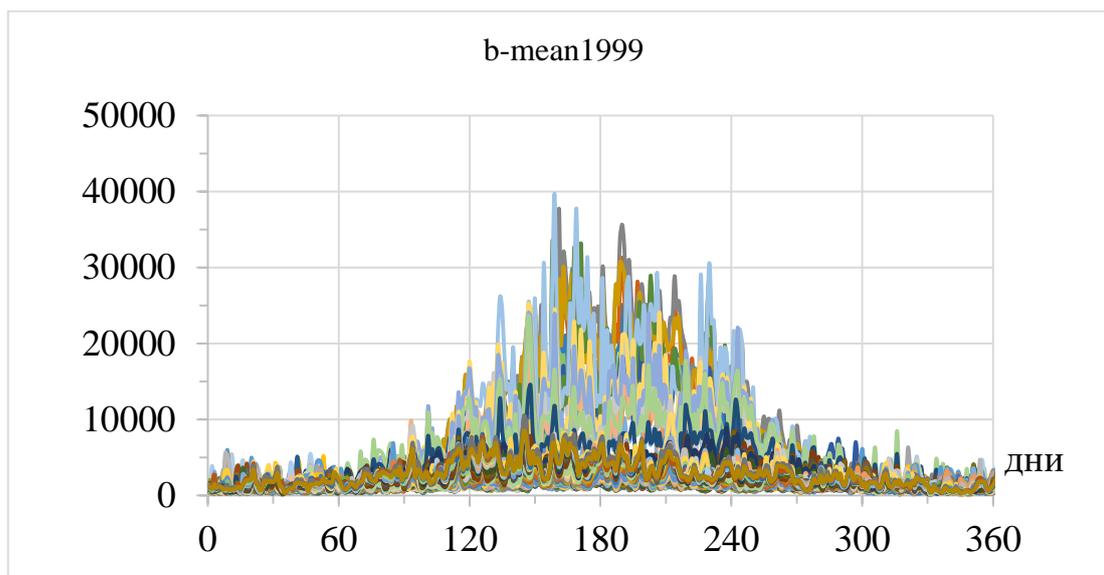
$$\frac{\partial p}{\partial t} = -\frac{\partial(\exp(i\omega t)A(x)p)}{\partial x} + \frac{1}{2} \frac{\partial^2(\exp(i\omega t)B(x)p)}{\partial x^2}, \quad (3)$$

где $p(t,x)$ – искомая плотность вероятности в момент t значения потока x , остальные обозначения приведены выше.

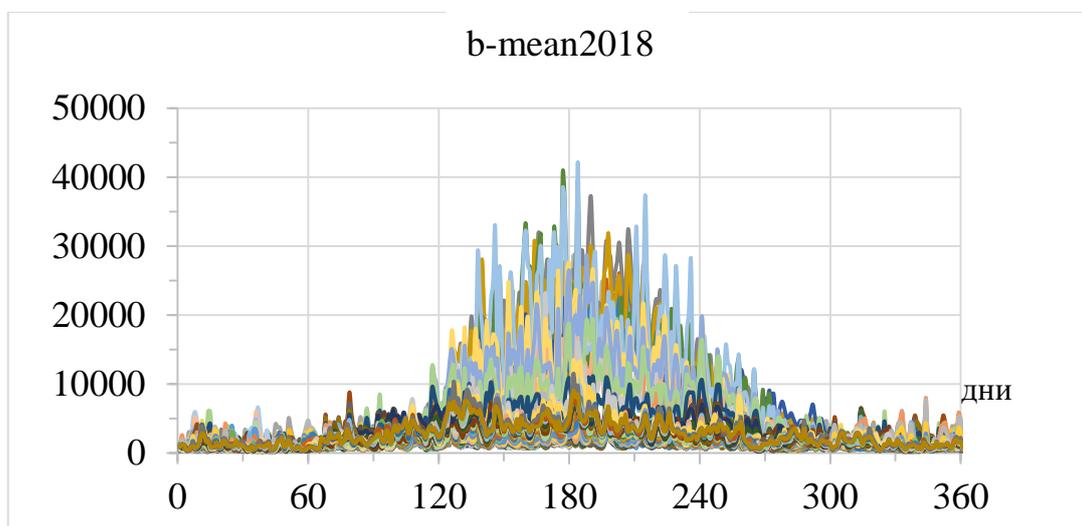
Уравнение (3) (уравнение Фоккера–Планка–Колмогорова [5, 6]) формально записано для комплексных коэффициентов, но физический смысл имеет только для вещественных значений [7]. Оно решается при граничных условиях $p(t, \pm\infty) = 0$ и начальном условии, если ставится задача Коши $p(0, x) = p_0(x)$, где $p_0(x)$ задано. Аналитически эта задача, вообще говоря, не решается, но численно ее решение не представляет особых трудностей. Однако, если потребовать, чтобы решение удовлетворяло определенным условиям на границе, например, чтобы $p(t, x) \cong \exp(-\lambda x^2)$ при $x \sim$ бесконечности, то обычными разностными методами это сделать непросто. Поэтому в работе используется модифицированный метод Галеркина, решающий эту проблему, детали которого будут составлять продолжение данных исследований.



(a)



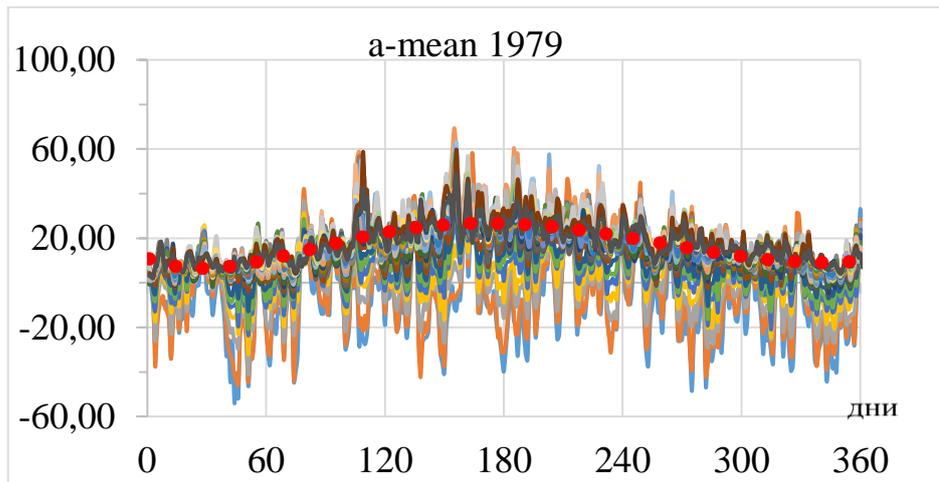
(б)



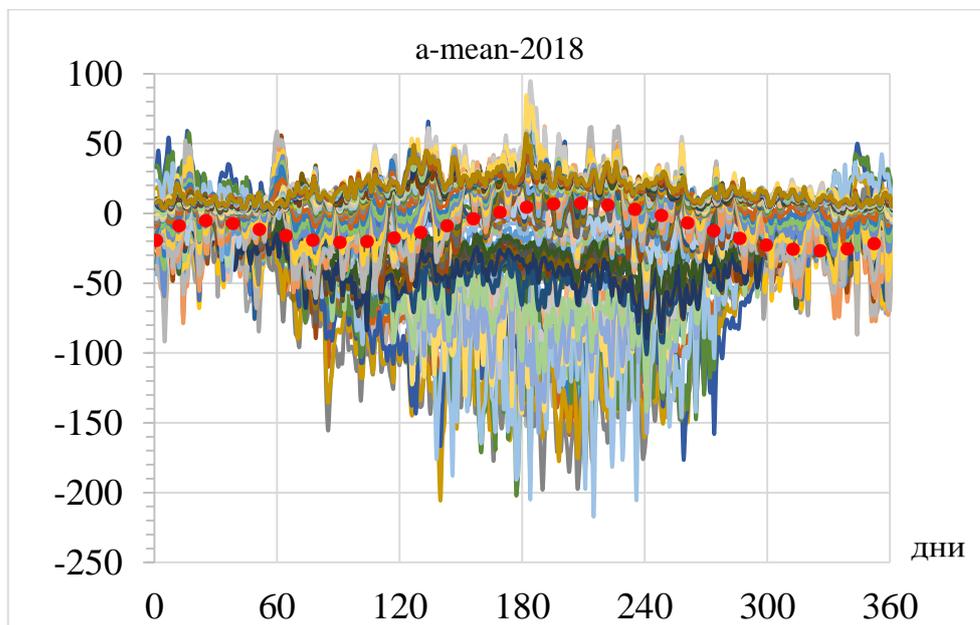
(в)

Рис. 3. Изменение средних значений коэффициента $b(t,x)$ с шагом 1 день за а) 1979; б) 1999; в) 2018 годы

Полученные зависимости для коэффициентов $a(t,x)$ и $b(t,x)$ были аппроксимированы аналитическими функциями, которые в дальнейшем используются для решения уравнения (1). Проанализированы тренды временного хода 40-летнего периода (Рис. 4 и 5).



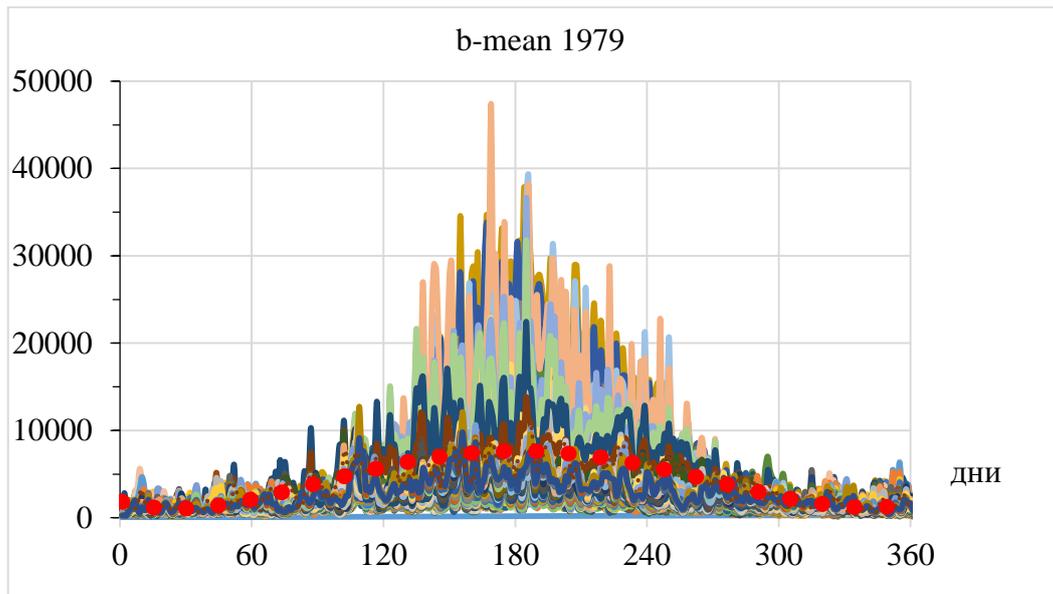
(а)



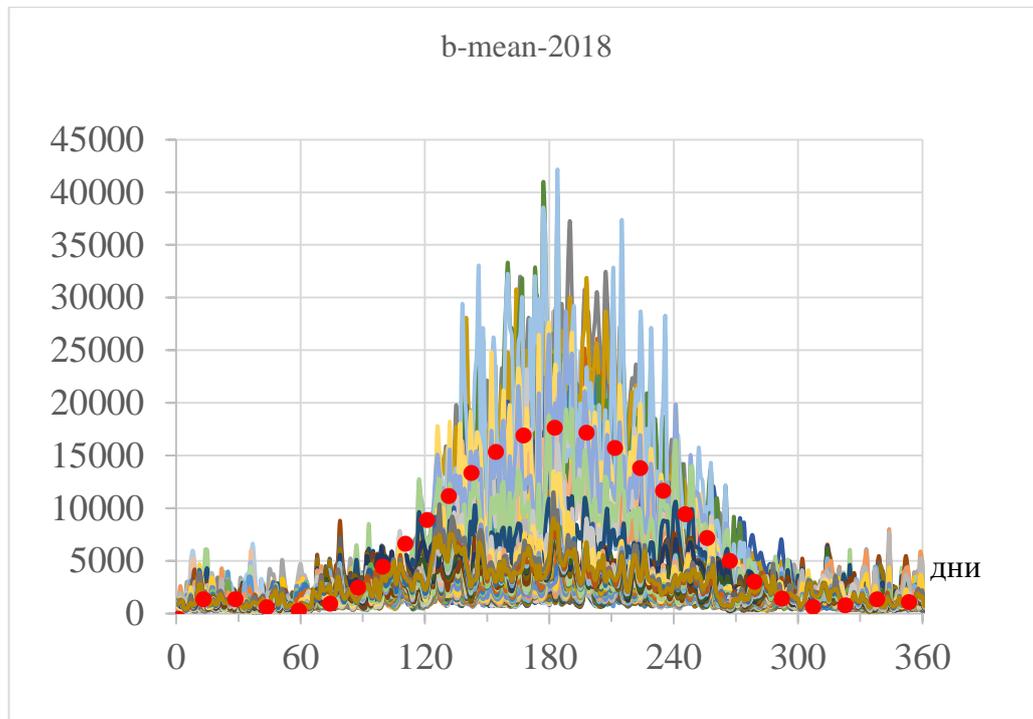
(б)

Рис. 4. Характеристики средних (снос) значений коэффициента $a(t,x)$, где красным пунктиром обозначены тренды: а) 1979 г. и б) 2018 г.

На Рис. 4(а, б) и 5(а, б) видно, что тренды значений коэффициентов $a(t,x)$ и $b(t,x)$ (красный пунктир) за 40 лет сохранились и носят более выраженный характер.



(а)



(б)

Рис. 5. Характеристики средних (диффузия) значений коэффициента $b(t,x)$, где красным пунктиром обозначены тренды: а) 1979 г. и б) 2018 г.

Дальнейшие исследования предполагают:

- *совместное исследование нескольких климатических характеристик (температуры и суммарного потока) методами вероятностного анализа;*
- *выявление регулярных и стохастических составляющих процессов*

для аналитической оценки аномальных явлений;

- *сравнение характеристик* различных периодов, в том числе аномальных.

БЛАГОДАРНОСТИ

Работа представлена в рамках выполнения темы НИР 0063-2019-0003 ФИЦ ИУ РАН и темы НИР 0128-2021-0002 ИО РАН.

Литература

1. *Belyaev K., Mikhaylov G., Salnikov A., Tuchkova N.* Seasonal and Decadal Variability of Atmosphere Pressure in Arctic, its Statistical and Temporal Analysis // CEUR Workshop Proceedings, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen (Aachen, Germany), 2020. V. 2784. P. 51–61.
<http://ceur-ws.org/Vol-2784/rpaper05.pdf>.
2. *Gavrikov A., Gulev S., Markina M., Tilinina N., Verezemskaya P., Barnier B., Dufour A., Zolina O., Zyulyaeva Y., Krinitskiy M., Okhlopkov I. Sokov A.* RAS-NAAD: 40-yr High-Resolution North Atlantic Atmospheric Hindcast for Multipurpose Applications (New Dataset for the Regional Mesoscale Studies in the Atmosphere and the Ocean) // Journal of Applied Meteorology and Climatology. 2020. V. 59, issue 5, P. 793–817.
<https://doi.org/10.1175/JAMC-D-19-0190.1>.
3. *Lemons Don S., Gythiel A.* Paul Langevin's 1908 paper "On the Theory of Brownian Motion" //Am. J. Phys. 1997. P. 65–111.
4. *Gythiel A.* Paul Langevin's 1908 paper "On the Theory of Brownian Motion" //Am. J. Phys. 1997. P. 65–111.
5. *Гухман И., Скороход А.* Введение в теорию случайных процессов. М.: Наука, 1965. 655 с.
6. *Risken H.* The Fokker–Planck Equation: Methods of Solutions and Applications. Springer. 1984. 452 p.
7. *Назаров А., Терпунов А.* Теория вероятностей и случайных процессов. Изд-во Томского государственного университета, 2010. 204 с.
8. *Беляев К.Л., Лаппо С.С., Музыченко А.С., Селеменов К.М.* Анализ статистических закономерностей временной изменчивости поверхностной температуры воды в Атлантическом и Тихом океанах // Сб. Статистические закономерности климатической изменчивости океанов. Ред. Лаппо С.С. Л: Гидрометиздат, 1988. С. 3–55.

STOCHASTIC ANALYSIS OF «OCEAN-ATMOSPHERE» HEAT FLUXES IN THE NORTH ATLANTIC

N. P. Tuchkova¹ [0000-0001-5357-9640], K. P. Belyaev² [0000-0003-2111-2709],

G. M. Mikhaylov³ [0000-0002-4535-7180]

¹⁻³*Dorodnicyn Computing Center FRC CSC of RAS, Vavilov str., 40, 11933, Moscow*

²*Shirshov Institute of Oceanology of RAS, Nahimovskiy pr., 36, 117218, Moscow*

¹natalia_tuchkova@mail.ru, ²kosbel55@gmail.com, ³gmickail@ccas.ru

Abstract

Observational data from the North Atlantic over 40 years of the NAAD project were analyzed. The total heat flux from the ocean to the atmosphere (and from the atmosphere to the ocean) was considered as the sum of latent and sensible heat. The coefficients of the stochastic differential equation representing the stochastic process were statistically determined from the original data set. Previously, the existence and uniqueness of a solution in the strong sense of the stochastic differential equation generated by the constructed diffusion process was proven when Kolmogorov's conditions were met. Numerical calculations on the Lomonosov-2 supercomputer at Moscow State University. M.V. Lomonosov.

Keywords: time series analysis, climatic seasonal cycle, maximum and minimum heat fluxes and temperature values within a climatic year

REFERENCES

1. Belyaev K., Mikhaylov G., Salnikov A., Tuchkova N. Seasonal and Decadal Variability of Atmosphere Pressure in Arctic, its Statistical and Temporal Analysis // CEUR Workshop Proceedings, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen (Aachen, Germany), 2020. V. 2784. P. 51–61.
<http://ceur-ws.org/Vol-2784/rpaper05.pdf>.

2. Gavrikov A., Gulev S., Markina M., Tilinina N., Verezemskaya P., Barnier B., Dufour A., Zolina O., Zyulyaeva Y., Krinitskiy M., Okhlopov I., Sokov A. RAS-NAAD: 40-yr High-Resolution North Atlantic Atmospheric Hindcast for Multipurpose Applications (New Dataset for the Regional Mesoscale Studies in the Atmosphere and the Ocean) // Journal of Applied Meteorology and Climatology. 2020, V. 59, issue 5. P. 793–817.

<https://doi.org/10.1175/JAMC-D-19-0190.1>.

3. *Lemons Don S., Gythiel A.* Paul Langevin's 1908 paper "On the Theory of Brownian Motion" //Am. J. Phys. 1997. P. 65–111.
4. *Gihman I., Skorohod A.* Vvedeniye v teoriyu sluchajnyh processov. M.: Nauka, 1965. 655 p.
5. *Risken H.* The Fokker–Planck Equation: Methods of Solutions and Applications. Springer. 1984. 452 p.
6. *Nazarov A., Terpunov A.* Teoriya veroyatnostej i sluchajnyh processov. Izdvo Tomskogo Gosuniversiteta. 2010. 204 p.
7. *Belyaev K., Lappo S., Muzychenko A., Selemenov K.* Analysis of statistical patterns of Sea Surface Temperature temporal variability in Atlantic and Pacific oceans. In "Statistical patterns of ocean climate variability. Ed. by Lappo S. S. L.: Hydrometizdat, 1988, p. 3–55.

СВЕДЕНИЯ ОБ АВТОРАХ



ТУЧКОВА Наталья Павловна – старший научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук, окончила ВМиК МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

Natalia Pavlovna TUCHKOVA – senior researcher of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree, graduated from CS Faculty of Lomonosov MSU. The expert in the field of algorithmic languages and information technologies.

email: natalia_tuchkova@mail.ru

ORCID: 0000-0001-5357-9640



БЕЛЯЕВ Константин Павлович – ведущий научный сотрудник Института океанологии им. П.П. Ширшова РАН и ФИЦ ИУ, доктор физ.-мат. наук, профессор кафедры теории вероятностей и статистики ВМиК МГУ им. М.В. Ломоносова. Сфера научных интересов – математическое моделирование и усвоение данных наблюдений, статистический анализ натуральных данных.

Konstantin Pavlovich BELYAEV – leading scientist of Shirshov Institute of Oceanology, Russian Academy of Science. Doctor of science, professor of Dept. of Applied Math and Cybernetics, Lomonosov Moscow State University. Research interests – math. modelling and data assimilation, statistical analysis of natural data.

email: kosbel55@gmail.com

ORCID: 0000-0003-2111-2709



МИХАЙЛОВ Гурій Михайлович – ведущий научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук. Сфера научных интересов – архитектура вычислительных систем и сетей, вычислительные и информационные технологии.

Gury Mickailovich Mikhaylov – leading scientist of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree. Research interests include architecture of computing systems and networks, computing and information technology.

email: gmickail@ccas.ru

ORCID: 0000-0002-4535-7180

Материал поступил в редакцию 18 декабря 2023 года

НА ПУТИ К СОЗДАНИЮ РАСПАРАЛЛЕЛИВАЮЩИХ КОМПИЛЯТОРОВ НА ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ С РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ

Б. Я. Штейнберг^[0000-0001-8146-0479]

Южный федеральный университет,
Институт математики, механики и компьютерных наук

byshtyaynberg@sfnedu.ru

Аннотация

Описаны условия создания оптимизирующих распараллеливающих компиляторов на вычислительные системы с распределённой памятью. Целевые вычислительные системы – это микросхемы типа «суперкомпьютер на кристалле». Приведены как оптимизирующие преобразования программ специфичные для систем с распределенной памятью, так и такие преобразования, которые нужны и для вычислительных систем с распределенной памятью, и для вычислительных систем с общей памятью. Обсуждены вопросы минимизации межпроцессорных пересылок при распараллеливании рекурсивной функции. Основной подход к созданию таких компиляторов – блочно-аффинные размещения данных в распределенной памяти с минимизацией межпроцессорных пересылок. Показано, что создавать распараллеливающие компиляторы на вычислительные системы с распределенной памятью следует на основе высокоуровневого внутреннего представления и с высокоуровневым выходным языком.

Ключевые слова: автоматизация распараллеливания, распределенная память, преобразования программ, размещение данных, пересылки данных

ВВЕДЕНИЕ

Во многих публикациях для создания параллельного кода для систем с распределенной памятью предложены полуавтоматические инструменты, в которых пользователь должен вызывать специальные функции или писать директивы компилятору. В работе [1] отмечено, что автоматическая компиляция последовательной программы для параллельной архитектуры с распределенной памятью

является очень сложной задачей, для которой нет в настоящее время эффективного решения. В этой же работе описана генерация параллельного кода, основанного на MPI (для кластера с мультипроцессорами). Об оптимизации размещения данных не говорится ничего. Использована распараллеливающая система Pluto, которая позволяет в пространстве итераций находить подмножества точек, допускающих параллельное выполнение. Ничего не говорится о локализации данных, которая необходима для эффективности на современных вычислительных архитектурах при параллельном выполнении относительно больших фрагментов кода [2], особенно при распределенной памяти.

В настоящей работе исследованы проблемы создания распараллеливающих компиляторов на вычислительные системы с распределенной памятью. Автоматическое распараллеливание развито только для вычислительных систем с общей памятью (см., например, [3–6]).

Для вычислительных систем с распределенной памятью (ВСРП) самой длительной операцией является пересылка данных между процессорными элементами (ПЭ). Для высокопроизводительных кластеров такие пересылки могут даже вызвать замедление, поскольку их длительность в десятки раз больше времени выполнения вычислительных операций. Для эффективного отображения на ВСРП программа должна удовлетворять таким жестким требованиям, что ускорение может быть достигнуто для узкого класса программ, и разработка компилятора нецелесообразна. Но в последнее время появляются многоядерные микросхемы, называемые «суперкомпьютер на кристалле», с десятками, сотнями и даже тысячами ядер, большинство из которых ориентированы на реализацию нейронных сетей [7–15]. Пересылка данных между процессорными ядрами одной микросхемы требует значительно меньше времени, чем на коммуникационной сети кластера. Это означает расширение множества эффективно распараллеливаемых программ и делает целесообразным разработку распараллеливающих компиляторов.

Много систем автоматизации создания параллельного кода для ВСРП используют дописывание прагм в последовательную программу без автоматизации предварительного преобразования [16–18].

В [19] отмечена перспективность циклических пересылок данных для микросхем близкого будущего (с тысячами ядер). В [20] приведено много задач линейной алгебры и математической физики, для параллельного решения которых на ВСРП использованы циклические пересылки. Метод размещения данных с перекрытиями [2] существенно ускоряет параллельные итерационные алгоритмы, уменьшая количество пересылок при укрупнении множеств пересылаемых данных. В [22] показан дополнительный эффект ускорения от размещений с перекрытиями для микросхем «суперкомпьютер на кристалле». В [23] и [24] описаны блочно-аффинные размещения данных в распределенной памяти.

В [21] рассмотрена задача распараллеливания программного цикла на ВСРП, где для минимизации межпроцессорных пересылок по тексту программы построен вспомогательный граф «операторы-переменные» и приведены примеры, основанные на OPC реализации [20].

Популярные оптимизирующие компиляторы (GCC, ICC, MS-Compiler, LLVM) распараллеливают программы только для вычислительных систем с общей памятью (ВСОП). Показано [25], что эти компиляторы имеют большой неиспользованный потенциал оптимизирующих преобразований. Эти результаты подтверждены в [26] и [27]. Можно полагать, что для микросхем supercomputer-on-a-chip этот потенциал неиспользуемых оптимизаций значительно больше.

Микросхемы «суперкомпьютер-на-кристалле» будут иметь большую производительность, чем многоядерные с общей памятью, и смогут выполнять новые классы программ с большим объемом вычислений. Отсутствие компиляторов на такие ВСРП сдерживает развитие новых высокопроизводительных систем.

В настоящей работе сформированы и обоснованы рекомендации к созданию оптимизирующих распараллеливающих компиляторов с последовательных программ языков высокого уровня для ВСРП, в частности, приведены оптимизирующие преобразования программ, которые должны использоваться в таком компиляторе. Рассмотрено также распараллеливание на ВСРП обхода дерева, которое задано рекурсивной функцией. Приведен пример распараллеливания гнезда циклов с условным оператором.

Подход, представленный ниже, отличается от других попыток автоматизировать отображение последовательных программ на ВСРП тем, что оптимальные

размещения массивов находятся среди блочно-аффинных размещений массивов [21], [23], которые, с одной стороны, могут быть описаны малым количеством параметров (пропорционально размерности массива), а, с другой стороны, покрывают размещения, широко используемые на практике. Расширение множества распараллеливаемых программ может быть достигнуто с помощью оптимизирующих преобразований программ, которые имеются в ОРС (Оптимизирующая распараллеливающая система) и в известных оптимизирующих компиляторах LLVM, GCC, ICC, MS-compiler.

1. МОДЕЛЬ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ С РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ

Будем рассматривать параллельную вычислительную систему с распределенной памятью (ВСП), состоящую из процессоров и модулей памяти, которые соединяются коммуникационной системой, по которой можно пересылать данные. Пересылки данных по коммуникационной сети требуют много времени и составляют основную задачу оптимизации программ на ВСП.

Существует множество конкретных вычислительных систем, которые удовлетворяют условиям описанной абстрактной модели. Многообразие таких вычислительных систем определяется разнообразием коммуникационных систем и вычислительных элементов и модулей памяти. ВСП отличаются не только количеством модулей памяти и вычислительных элементов, но и их связями посредством коммуникационной сети, которые по длительности могут быть несимметричными. Распределенная память входит в состав вычислительных систем с программируемой архитектурой [28]. Как только будет разработан новый высокопроизводительный компьютер относительно небольших габаритов, сразу появится желание для некоторых важных задач объединить несколько таких компьютеров сетью – и появится новая ВСП. Еще много архитектур будет придумано.

Многообразие архитектур ВСП создает следующее требование к архитектуре преобразователей программ: должна быть система преобразований программ, которые полезны широкому множеству ВСП, и должны быть библиотеки преобразований, ориентированные на конкретную целевую архитектуру. Эти требования аналогично тому, как устроены семейства компиляторов GCC и LLVM: есть библиотека оптимизирующих преобразований программ в промежуточном

представлении (middle end) и есть генераторы кода на целевую архитектуру (back end), которые тоже содержат оптимизирующие преобразования.

2. О ПРЕОБРАЗОВАНИЯХ ПРОГРАММ ДЛЯ ВСРП, КОТОРЫЕ НЕ НУЖНЫ ДЛЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ С ОБЩЕЙ ПАМЯТЬЮ

Отображение последовательных программ на ВСРП требует разработки дополнительных функций, которых нет в вычислительных системах с общей памятью (ВСОП), например: построение вспомогательного графа «операторы-переменные» для поиска минимального множества пересылок, генерация межпроцессорных пересылок, группировка пересылок, поиск оптимального размещения данных и кратного оптимального размещения данных в распределенной памяти с перекрытиями.

Оптимизирующие преобразования для ВСРП направлены, в первую очередь, на минимизацию пересылок между модулями системы. Можно выделить следующие функции и преобразования:

- Размещение данных в распределенной памяти.
- Поиск оптимального множества циклических пересылок с помощью ГОП (граф «операторы-переменные»).
- Транспонирование матриц, размещенных в распределенной памяти, и его обобщение на многомерные массивы.
- Размещение массивов данных кратными с перекрытиями.
- Группировка пересылок.
- Определение оптимального количества ПЭ.
- Оптимизация преобразований многомерных массивов: транспонирование и скашивание.

К преобразованиям, специфическим для ВСРП, можно отнести распознавание и вызов пересылок данных для многомерных массивов, таких как транспонирование и скашивание.

Транспонирование матриц используется и в библиотеках прикладных программ для обычных процессоров, поскольку от расположения элементов матрицы в оперативной памяти зависит скорость их считывания. Транспонирование может быть описано простой программой

```
for (int i = 0; i < n; i++) {  
    for (int j = 0; j < n; j++) {  
        Y(i,j) = X(j,i);  
    }  
}
```

В ВСПП транспонирование матрицы может быть полезно, если в одной части программы матрица должна быть размещена в распределенной памяти по строкам, а в другой – по столбцам. Если размерность матрицы равна количеству ПЭ и коммуникационная система поддерживает циклические пересылки данных, то транспонирование сводится к одновременным пересылкам диагональных элементов.

При скошенной форме хранения в ПЭ хранятся косые диагонали матрицы (диагонали, перпендикулярные к главной) [20]. Скошенная форма хранения позволяет обращаться одновременно как к элементам любой строки матрицы-клетки, так и к элементам любого столбца. Алгоритм приведения матрицы к скошенной форме может выглядеть следующим образом.

```
for (int i = 0; i < n; i++) {  
    for (int j = 0; j < n; j++) {  
        Y(i,j) = X((j+i) mod n ,i);  
    }  
}
```

3. ОПТИМИЗИРУЮЩИЕ ПРЕОБРАЗОВАНИЯ ПРОГРАММ, КОТОРЫЕ ПОЛЕЗНЫ ОДНОВРЕМЕННО И ДЛЯ ВСПП, И ДЛЯ ВСОП

В данном разделе рассмотрены предварительные преобразования последовательных программ перед отображением на вычислительную систему, которая может иметь как распределенную память, так и общую.

Оптимизировать следует участки кода, требующие больших объемов вычислений. Такими участками являются структуры вложенных циклов или рекурсивные функции. Обычные компиляторы LLVM, GCC, ICC, MS-compiler хорошо оптимизируют базовые блоки и самые глубоко вложенные циклы [30]. Такая оптимизация ориентирована на использование регистров, АЛУ и векторизацию. Но в последнее время в архитектуре компьютеров появилась сложная структура памяти даже в ВСОП (распределенная память ВСПП и так сегодня непосильна для

компиляторов). Для оптимизации использования сложной структуры памяти появилась потребность в преобразованиях циклов, которые могут содержать другие циклы (в первую очередь, тайлинг и скошенный тайлинг) [31], [30], а в преобразованиях данных – блочные размещения матриц в оперативной памяти [30], блочно-аффинные и размещения массивов и размещения с перекрытиями в распределенной памяти [30] и т. п.

К преобразованиям циклов следует отнести слияние, разбиение, гнездование, развертку, расщепление и др. Это преобразования, которые традиционно выполняются для самых глубоко вложенных циклов (innermost loop), но условия эффективного применения этих преобразований недостаточно разработаны к не самым глубоко вложенным циклам, и эти преобразования компиляторами не выполняются [32]. Есть преобразования, описанные в литературе, которые не выполняются в компиляторах, такие как ретайминг (сдвиги операторов между итерациями циклов) [30].

Диагональное расщепление двойного гнезда циклов позволяет избавиться от операции взятия остатка от деления при вычислениях с целыми числами. Оно встречается для быстрого вычисления сверток на циклической группе при восстановлении изображений [33], в теории кодирования и других приложениях. В компиляторах оно пока не реализуется. На микросхемах с распределенной локальной памятью названное преобразование актуально вместе с его применениями.

Не поддерживается компиляторами оптимизация обхода дерева вариантов, который возникает при рекурсивных вызовах функций, например, для многих задач, решаемых методом ветвей и границ. Нет инструментов, поддерживающих параллельный обход ветвей такого дерева [30].

Главная проблема современных оптимизирующих компиляторов на ВСОП – выбор оптимальной цепочки преобразований [30]. Для компиляторов на ВСПП эта проблема будет еще острее. Распараллеливание дерева обхода таких цепочек может быть полезно.

4. ПАРАЛЛЕЛЬНЫЙ ОБХОД ДЕРЕВА ПРИ РЕКУРСИВНОМ ВЫЗОВЕ ФУНКЦИИ

Рекурсивный вызов функции во многих случаях порождает процедуру обхода дерева. С другой стороны, обход дерева можно описывать рекурсивной функцией. Вызовы рекурсивных функций, как и программные циклы, во многих случаях требуют больших объемов вычислений и нуждаются в ускорениях. Обход дерева возникает в прикладных переборных задачах, многие из которых являются NP-трудными (например, известные задача коммивояжера или задача о ранце). В частности, обход дерева является составной частью метода ветвей и границ.

Многие переборные задачи не решаются точными алгоритмами, поскольку объемы вычислений не позволяют найти такое решение за приемлемые расходы времени и средств на современных компьютерах. Но приходит новое поколение вычислительных систем, и многие задачи, которые раньше не рассматривались как потенциально решаемые, теперь могут найти свое решение. Кроме того, на основе точных алгоритмов обхода дерева могут строиться более быстрые эвристические алгоритмы.

Естественный и обычно используемый метод распараллеливания обхода дерева – параллельное выполнение обхода разных ветвей [34], [35].

Рассмотрим оптимизацию обхода дерева рекурсивной функцией.

Будем рассматривать частичный инлайнинг рекурсивной функции. Вызов функции при трансляции в ассемблер создает дополнительные операции, такие как присваивания формальным параметрам их конкретных значений, после чего возможны упрощения арифметических выражений, и другие. Поэтому инлайнинг (подстановка вместо вызова функции тела этой функции) уменьшает количество операций и, как правило, ускоряет код. Если функция вызывается очень много раз, то выполнение инлайнинга для всех вызовов сильно увеличивает объем кода, что может приводить к замедлению программы.

У рекурсивной функции инлайнинг не исключает операцию вызова функции, и можно говорить только о частичном инлайнинге.

Рекурсивная функция создает обход дерева, если в ней есть по крайней мере два рекурсивных вызова, которые не лежат на альтернативных ветках условного оператора. Рассмотрим псевдокод такой функции, имеющей для простоты изложения два рекурсивных вызова.

Function F1

```
{B1;
```

```
Call F1;
```

```
B2;
```

```
Call F1;
```

```
B3;}
```

В этой функции можно сделать частичный инлайнинг обоих вызовов и получить новую функцию

Function F2

```
{B1;
```

```
    {B1;
```

```
    Call F1;
```

```
    B2;
```

```
    Call F1;
```

```
    B3;}
```

```
B2;
```

```
    {B1;
```

```
    Call F1;
```

```
    B2;
```

```
    Call F1;
```

```
    B3;}
```

```
B3;}
```

Таким образом, после частичного инлайнинга получено четыре вызова исходной функции F1. Можно повторить частичный инлайнинг несколько раз и получить много вызовов функции F1.

Выполнение исходной функции можно распараллелить, параллельно выполняя разные вызовы функции F1. Каждый вызов функции F1 соответствует обходу некоторой ветви дерева вычислений функции исходной программы.

Чем больше раз выполняется частичный инлайнинг, тем больше создается вызовов функций и, следовательно, параллельно выполняемых процессов.

Во многих алгоритмах, основанных на обходе дерева, особенно с использованием метода ветвей и границ, время выполнения разных ветвей дерева может очень сильно различаться. Иногда время обхода одной ветви дерева может быть больше времени обхода остальных вместе взятых. По этой причине параллельное выполнение разных ветвей дерева может оказаться плохо сбалансированным.

Если вызовов функций (ветвей дерева вычислений) больше, чем вычислительных устройств (процессоров, процессорных элементов, процессорных ядер, ...), то некоторым устройствам придется выполнять по несколько вызовов. Но заранее не известно, какой вызов функции F1 как долго будет обрабатываться.

Для улучшения балансировки нагрузки вычислительных устройств удобно использовать технологии параллельного программирования типа TPL или Intel TBB, в которых создается очередь заданий (вызовов функций), и вычислительные устройства получают новые задания из очереди, как только они освобождаются от выполнения предыдущего.

Чем большее количество мелких заданий в очереди, тем лучше может быть сбалансирована нагрузка вычислительных устройств. Но, с другой стороны, увеличиваются накладные расходы с запуском на выполнение очередного задания из очереди. Запуск задания может быть связан не только с командами запуска, но и с подготовкой данных, особенно при распределенной памяти.

Будем минимизировать время на пересылки заданий между ПЭ. Будет полагать, что ПЭ соединяются шиной или кольцевой сетью. Пусть в некоторый момент работы алгоритма каждый ПЭ имеет некоторое множество заданий и освободился один ПЭ (количество его заданий равно нулю), которому требуется переслать задание. Конечно, быстрее всего переслать задание из соседнего ПЭ. Но шина (или кольцевая шина) позволяет в одно и то же время делать несколько пересылок в одном направлении. Этим самым можно подготовиться к минимизации времени следующих пересылок.

Наилучшим представляется распределение количества заданий, при котором в каждом ПЭ одинаковое количество заданий. Таким образом, будем делать

пересылки, при которых минимизируется среднее квадратичное отклонение от среднего количества заданий в ПЭ.

Пример

Номер ПЭ	0	1	2	3	4	5	6	7
К-во заданий	4	4	2	3	0	2	4	1

Среднее количество заданий в ПЭ равно 2.5.

Разумной (оптимальной) представляется следующая пересылка заданий слева направо:

ПЭ0→ПЭ1; ПЭ1→ПЭ2; ПЭ3→ПЭ4; ПЭ6→ПЭ7.

После такой пересылки распределение заданий будет таким

Номер ПЭ	0	1	2	3	4	5	6	7
К-во заданий	3	3	2	2	1	2	3	2

Среднее квадратичное отклонение от среднего количества заданий равно

$$(1.5^2+1.5^2+0.5^2+0.5^2+1.5^2+0.5^2+0.5^2+0.5^2)/8=1.0.$$

Приведем алгоритм для поиска оптимальной пересылки заданий по шине.

Задача алгоритма – не только загрузить ПЭ, у которого нет заданий, но и в это же время (этой же пересылкой) уменьшить среднее квадратичное отклонение от среднего количества заданий. Это приведет к уменьшению времени выполнения при предположении, что время выполнения всех заданий одинаково.

Будем отдельно искать оптимальный алгоритм пересылки данных по шине слева направо и справа налево, затем выберем из них лучший. Ясно, что достаточно описать только один алгоритм, например, поиска оптимальной пересылки слева направо.

Пусть количество ПЭ равно n , и они занумерованы от 0 до $(n-1)$. Обозначим b_j количество заданий в ПЭ с номером j . Ясно, что $b_j \geq 0$. Обозначим через x_j количество заданий, пересылаемых из ПЭ с номером j в ПЭ с номером $(j+1)$. Величина x_j принимает значения 0 или 1. В результате работы алгоритма определяются все значения x_j .

Алгоритм (жадный)

Если у каждого ПЭ количество заданий не больше 1, то конец работы алгоритма (т. е. ПЭ, у которых есть задания, завершают их выполнение, пересылки заданий нецелесообразны).

Полагаем $x_{n-1}=0$.

Просматриваем все ПЭ с номерами j справа налево, начиная с $(n-2)$ -го до 0-го. Если $b_j > b_{j+1} - x_{j+1}$, то $x_j=1$, иначе $x_j=0$.

В результате действия данного алгоритма не увеличивается количество заданий каждого ПЭ. Если хотя бы у одного ПЭ количество заданий больше 1 при том, что у какого то ПЭ заданий 0, то данный алгоритм найдет пересылку либо слева направо, либо справа налево, которая уменьшит среднее квадратичное отклонение количества заданий от среднего количества заданий.

5. О СОЗДАНИИ АРХИТЕКТУРЫ РАСПАРАЛЛЕЛИВАЮЩЕЙ СИСТЕМЫ НА ВСРП

Даже для ВСОП распараллеливающий компилятор – это программный продукт, большой по объему кода, обладающий сложной логикой, требующий очень высокой квалификации разработчиков и значительного времени разработки. По этой причине при разработке компилятора на новый процессор используются большие части кода компиляторов, уже созданных для прежних вычислительных систем (например, оптимизирующие преобразования LLVM). Оригинальными являются генератор кода и некоторые комбинации существующих преобразований. Такой подход по экономическим причинам неизбежен (с дополнительными усложнениями) и для компиляторов на ВСРП.

Специфические преобразования программ на ВСРП предполагают выделение их в отдельную библиотеку. Поскольку для многих архитектур ВСРП выход компилятора должен содержать участки кода, выполняемые на ВСОП, результатом таких преобразований должен быть, скорее всего, высокоуровневый язык, к которому можно применить существующий компилятор на ВСОП (например, ICC или C-to-CUDA). Поскольку самый быстрый код создается программами языков C и ФОРТРАН, то на эти языки и должна быть рассчитана распараллеливающая система на ВСРП. Такая система будет совместима со всеми компиляторами. Даже для ВСОП преобразования гнезд циклов удобнее выполнять с выходным высокоуровневым языком [36].

Многообразие архитектур ВСРП предполагает создание, в первую очередь, преобразований программ, полезных для многих ВСРП: размещение данных в

распределенной памяти (блочно-аффинные), алгоритмы поиска минимального множества циклических пересылок и рассылок данных (broadcast).

Внутреннее (промежуточное) представление обсуждаемой распараллеливающей системы желательно делать высокоуровневым. Высокоуровневое внутреннее представление позволит создавать выходной код, более близкий (узнаваемый) к исходному коду разработчика последовательной программы. К такому коду легче применять диалоговый режим уточнения зависимостей [29]. Известные распараллеливающие системы, выдающие код на высокоуровневом языке, имеют, как правило, высокоуровневое внутреннее представление: ROSE, SUIF, OPS, PLUTO.

6. ОПТИМИЗИРУЮЩИЕ КОМПИЛЯТОРЫ VS БИБЛИОТЕК

Разработчики процессоров ВСОП в качестве системного ПО выпускают компиляторы и библиотеки. Для создания высокопроизводительных приложений следует использовать библиотеки. Но они разрабатываются не для всех приложений. При выпуске новой микросхемы необходимо все библиотеки переписывать.

Для ВСРП создание библиотек требует существенно больших усилий, чем для ВСОП, поскольку исходные данные для библиотечных функций в разных случаях могут быть размещены по-разному. Рассмотрим, например, задачу перемножения матриц $C=A*B$. Если размерность исходных матриц равна количеству ПЭ, то в случаях, рассматриваемых на практике, каждая матрица может размещаться либо «по строкам», либо «по столбцам», либо «в скошенной форме». В этом случае нужно 9 библиотечных функций. Но, если размерности матриц больше, чем количество ПЭ, то матрицы могут размещаться полосами строк или столбцов, что увеличивает количество библиотечных функций.

Если ВСРП используется как ускоритель некоторого класса приложений, то библиотеки могут иметь преимущество по сравнению с компилятором. Если ВСРП предполагается использовать для переноса программ с других вычислительных систем или для разработки новых приложений, то использование компилятора более предпочтительно.

Есть еще одна функция для использования высокоуровневого преобразователя программ – проектирование новых вычислительных систем. Действительно,

эти системы при проектировании подстраиваются под некоторые бенчмарки. Бенчмарки – это конкретные программные реализации некоторых полезных алгоритмов. Но алгоритмы могут быть реализованы не единственным образом. Высокоуровневый преобразователь программ может представить семейство программ, эквивалентных исходной (бенчмарку). Проектируемая вычислительная система может ориентироваться на любой экземпляр семейства.

7. ЗАКЛЮЧЕНИЕ

В настоящей статье приведены основные элементы проекта системы преобразования программ для создания распараллеливающих компиляторов на вычислительные системы с распределенной памятью.

Существует много приложений, для которых не видно предела в повышении производительности вычислительных систем: прогноз погоды и изменений климата, отслеживание угроз столкновения метеоритов с Землей, планирование экономики, создание новых лекарств, решение задач метагеномики и др. Потребность в новых высокопроизводительных программно-аппаратных комплексах будет у общества еще многие годы. Дорогой проект создания высокоуровневого преобразователя программ для ВСРП является вполне экономически целесообразным.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00671, <https://rscf.ru/project/22-21-00671/>.

СПИСОК ЛИТЕРАТУРЫ

1. *Bondhugula U.* Automatic distributed-memory parallelization and codegeneration using the polyhedral framework, Technical report, ISc-CSA-TR-2011-3, 2011, 10 p.
2. *Ammaev S.G., Gervich L.R., Steinberg B.Y.* Combining parallelization with overlaps and optimization of cache memory usage. PaCT 2017: Parallel Computing Technologies, Lecture Notes in Computer Science. Vol. 10421. P. 257–264.
3. Векторизация программ // Векторизация программ: теория, методы, реализация / Сборник переводов статей. М.: Мир, 1991. С. 246–267.

4. *Moldovanova O.V., Kurnosov M.G.* Auto-Vectorization of Loops on Intel 64 and Intel Xeon Phi: Analysis and Evaluation International Conference on Parallel Computing Technologies. PaCT 2017: Parallel Computing Technologies, Lecture Notes in Computer Science. Vol. 10421. P. 143–150.

5. Nvidia compilers. URL: <https://developer.nvidia.com/hpc-compiler>

6. *Peng Di, Ding Ye, Yu Su, Yulei Sui, Jingling Xue.* Automatic Parallelization of Tiled Loop Nests with Enhanced Fine-Grained Parallelism on GPUs. 2012. 41st International Conference on Parallel Computing.

7. SoC Esperanto. URL: <https://www.esperanto.ai/technology/>

8. Процессор HTЦ «Модуль».

URL: https://www.cnews.ru/news/top/2019-03-06_svet_uvidel_moshchnejshij_rossijskij_nejroprotessor (дата обр. 26.03.2022).

9. *Peckham O. SambaNova.* Launches Second-Gen DataScale System.

URL: <https://www.hpcwire.com/2022/09/14/sambanova-launches-second-gen-datascalesystem/>.

10. *Елизаров Г.С., Конотопцев В.Н., Корнеев В.В.* Специализированные большие интегральные схемы для реализации нейросетевого вывода. XXII международная конференция «Харитоновские тематические научные чтения». Суперкомпьютерное моделирование и искусственный интеллект: труды / Редактор Р.М. Шагалиев. Саров: ФГУП «РФЯЦ-ВНИЭФ», 2022. С. 181–184.

11. *Корнеев В.В.* Направления повышения производительности нейросетевых вычислений // Программная инженерия. 2020. Т. 11, № 1. С. 21–25.

12. *Yen I.E., Xiao Zh., Xu D.* S4: a High-sparsity, High-performance AI Accelerator // arXiv:2207.08006v1 [cs.AR] 16 Jul 2022.

13. *Gale T., Elsen E., Hooker S.* The state of sparsity in deep neural networks // arXiv preprint arXiv:1902.09574, 2019.

14. Intelligence Processing Unit. URL: <https://www.graphcore.ai/products/ipu>.

15. *Jia Zh., Tillman B., Maggioni M., Scarpazza D.P.* Dissecting the Graphcore IPU Architecture via Microbenchmarking // Technical Report. December 7, 2019. arXiv:1912.03413v1 [cs.DC] 7 Dec 2019. 91 p.

16. DVM-система разработки параллельных программ.

URL: <http://dvm-system.org/ru/about/>

17. *Kataev N., Kolganov A.* Additional Parallelization of Existing MPI Programs Using SAPFOR. In: Malyshkin V. (Ed.) *Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science.* 2021. Vol. 12942. Springer, Cham.

URL: https://doi.org/10.1007/978-3-030-86359-3_4

18. *Kwon D., Han S., Kim H.* MPI backend for an automatic parallelizing compiler // *Proceedings Fourth International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN'99).* 06.1999. P. 152–157.

19. *Корнеев В.В.* Параллельное программирование // *Программная инженерия.* 2022. Т. 13, № 1. С. 3–16.

20. *Прангишвили И.В., Виленкин С.Я., Медведев И.Л.* Параллельные вычислительные системы с общим управлением. М.: Энергоатомиздат, 1983. 312 с.

21. *Krivosheev N.M., Steinberg B.Ya.* Algorithm for searching minimum inter-node data transfers // *Procedia Computer Science, 10th International Young Scientist Conference on Computational Science, YSC 2021, 1–3 July 2021.* P. 306–313.

22. *Gervich L.R., Steinberg B.Ya.* Automation of the Application of Data Distribution with Overlapping in Distributed Memory // *Bulletin of the South Ural State University. Ser. Mathematical Modeling, Programming & Computer Software (Bulletin SUSU MMCS).* 2023. Vol. 16, no. 1. P. 59–68.

23. *Штейнберг Б.Я.* Блочно-аффинные размещения данных в параллельной памяти // *Информационные технологии.* 2010. №6. С. 36–41.

24. *Штейнберг Б.Я.* Оптимизация размещения данных в параллельной памяти. Ростов-на-Дону, Изд-во Южного федерального университета, 2010. 255 с.

25. *Gong Z., Chen Z., Szaday Z., Wong D., Sura Z., Watkinson N., Maleki S., Padua D., Veidenbaum A., Nicolau A.* An empirical study of the effect of source-level loop transformations on compiler stability // *Proceedings of the ACM on Programming Languages.* 11.2018. P. 1–29.

26. *Steinberg B.Ya., Steinberg O.B., Oganesyanyan P.A., Vasilenko A.A., Veselovskiy V.V., Zhivykh N.A.* Fast Solvers for Systems of Linear Equations with Block-Band Matrices // *East Asian Journal on Applied Mathematics.* 2023. Vol. 13, No. 1. P. 47–58.

27. *Vasilenko A., Veselovskiy V., Metelitsa E., Zhivykh N., Steinberg B., Steinberg O.* Precompiler for the ACELAN-COMPOS Package Solvers // In: Malyshkin V. (Ed.)

Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science. 2021. Vol. 12942. P. 103–116. Springer, Cham.

URL: https://doi.org/10.1007/978-3-030-86359-3_8

28. *Dordopulo A.I., Levin I.I., Gudkov V.A., Gulenok A.A.* High-Level Synthesis of Scalable Solutions from C-Programs for Reconfigurable Computer Systems // In: Malyshkin V. (Ed.) Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science. 2021. Vol. 12942. Springer, Cham.

URL: https://doi.org/10.1007/978-3-030-86359-3_7

29. *Штейнберг Б.Я.* Блочно рекуррентное размещение матрицы для параллельного выполнения алгоритма Флойда // Известия ВУЗов. Северокавказский регион. Естественные науки. 2010. №5. С. 31–33.

30. *Штейнберг Б.Я., Штейнберг О.Б.* Преобразования программ – фундаментальная основа создания оптимизирующих распараллеливающих компиляторов // Программные системы: теория и приложения. 2021. Т. 12, № 1(48). С. 21–113. URL: http://psta.psisras.ru/read/psta2021_1_21-113.pdf

31. *Wolfe M.* More Iteration Space Tiling // Supercomputing. Reno, 1989. P. 655–664.

32. *Штейнберг Б.Я., Штейнберг О.Б., Василенко А.А.* Слияние циклов для локализации данных // Программные системы. Теория и приложения. 2020. Т. 11, №3. С. 17–31. URL: <https://doi.org/10.25209/2079-3316-2020-11-3-17-31>

33. *Козак А.В., Штейнберг Б.Я., Штейнберг О.Б.* Алгоритм восстановления смазанного изображения, полученного вращающейся под углом к горизонту камерой // Компьютерная оптика. 2020. Т. 44, № 2. С. 229–235.

34. *Burkhovetskiy V.V., Steinberg B.Ya.* Parallelizing an Exact Algorithm for the Traveling Salesman Problem // Procedia Computer Science, 6-th International Young Scientist Conference on Computational Science, YSC 2017, Procedia Computer Science. 2017. Vol. 119. P. 97–102.

URL: <http://authors.elsevier.com/sd/article/S187705091732375X>

35. *Бурховецкий В.В., Штейнберг Б.Я.* Стратегия использования крупных заданий при параллельном обходе дерева // Языки программирования и компиля-

торы. Труды Всероссийской научной конференции памяти А.Л. Фуксмана. 3–5 апреля 2017, Южный федеральный университет, г. Ростов-на-Дону: Изд-во Южного федерального университета. 2017. С. 66–70.

36. *Zhiyuan Li, Yonghong Song*. Automatic Tiling of Iterative Stencil Loops // ACM Transactions on Programming Languages and Systems. 2004. Vol. 26, No. 6. P. 975–1028.

37. *Gervich L.R., Guda S.A., Dubrov D.V., Ibragimov R.A., Metelitsa E.A., Mikhailuts Y.M., Paterikin A.E., Petrenko V.V., Skapenko I.R., Steinberg B.Ya., Steinberg O.B., Yakovlev V.A., Yurushkin M.V.* How OPS (Optimizing Parallelizing System) May be Useful for Clang // CEE-SECR '2017, October 20–21, 2017, St.-Peterburg, Russian Federation. Proceedings of the 13th Central & Eastern European Software Engineering Conference in Russia ACM New York, NY, USA. 2017.

URL: <https://dl.acm.org/citation.cfm?id=3166094&picked=prox>

ON THE WAY TO CREATING PARALLELIZING COMPILERS FOR COMPUTING SYSTEMS WITH DISTRIBUTED MEMORY

B. Ya. Steinberg^[0000-0001-8146-0479]

Southern federal university

byshtyaynberg@sfnedu.ru

Abstract

The conditions for creating optimizing parallelizing compilers for computing systems with distributed memory are described. Target computing systems are microcircuits of the “supercomputer on a chip” type. Both optimizing program transformations specific to systems with distributed memory and those transformations that are needed both for computing systems with distributed memory and for computing systems with shared memory are presented. The issues of minimizing interprocessor transfers when parallelizing a recursive function are discussed. The main approach to creating such compilers is block-affine data placement in distributed memory with minimization of inter-processor transfers. It is shown that parallelizing compilers for computing systems with distributed memory should be created on the basis of a high-level internal representation and a high-level output language.

Keywords: *automatic parallelization, distributed memory, program transformation, data distribution, data interchange*

REFERENCES

1. *Bondhugula U.* Automatic distributed-memory parallelization and codegeneration using the polyhedral framework, Technical report, ISc-CSA-TR-2011-3, 2011, 10 p.
2. *Ammaev S.G., Gervich L.R., Steinberg B.Y.* Combining parallelization with overlaps and optimization of cache memory usage. PaCT 2017: Parallel Computing Technologies, Lecture Notes in Computer Science. Vol. 10421. P. 257–264.
3. *Vektorizaciya programm // Vektorizaciya programm: teoriya, metody, realizaciya. / Sbornik perevodov statej. M.: Mir, 1991. S. 246–267.*
4. *Moldovanova O.V., Kurnosov M.G.* Auto-Vectorization of Loops on Intel 64 and Intel Xeon Phi: Analysis and Evaluation International Conference on Parallel Computing Technologies. PaCT 2017: Parallel Computing Technologies, Lecture Notes in Computer Science. Vol. 10421. P. 143–150.
5. Nvidia compilers. URL: <https://developer.nvidia.com/hpc-compiler>
6. *Peng Di, Ding Ye, Yu Su, Yulei Sui, Jingling Xue.* Automatic Parallelization of Tiled Loop Nests with Enhanced Fine-Grained Parallelism on GPUs. 2012. 41st International Conference on Parallel Computing.
7. SoC Esperanto. URL: <https://www.esperanto.ai/technology/>
8. Processor NTC “Modul”.
URL: https://www.cnews.ru/news/top/2019-03-06_svet_uvidel_moshchnejshij_rossijskij_nejroprotsessor
9. *Peckham O. SambaNova.* Launches Second-Gen DataScale System.
URL: <https://www.hpcwire.com/2022/09/14/sambanova-launches-second-gen-datascalesystem/>.
10. *Elizarov G.S., Konoptsev V.N., Korneev V.V.* Specialized large integrated circuits for the implementation of neural network inference // XXII International conference "Kharitonov thematic scientific readings "Supercomputer modeling and Artificial-Intelligence": proceedings / Edited by R.M. Shagaliev. Sarov: FSUE "RFSC-VNIIEF", 2022. P. 181–184.

11. *Korneev V.V.* Approaches to improving the performance of neural network computing // *Programmnyaya inzheneriya*. 2020. Vol. 11. No. 1. P. 21–25.
URL: <https://doi.org/10.17587/prin.11.21-25>.
 12. *Yen I.E., Xiao Zh., Xu D.* S4: a High-sparsity, High-performance AI Accelerator // arXiv:2207.08006v1 [cs.AR] 16 Jul 2022.
 13. *Gale T., Elsen E., Hooker S.* The state of sparsity in deep neural networks // arXiv preprint arXiv:1902.09574, 2019.
 14. Intelligence Processing Unit.
URL: <https://www.graphcore.ai/products/ipu>.
 15. *Jia Zh., Tillman B., Maggioni M., Scarpazza D.P.* Dissecting the Graphcore IPU Architecture via Microbenchmarking // Technical Report. December 7, 2019. arXiv:1912.03413v1 [cs.DC] 7 Dec 2019. 91 p.
 16. DVM-sistema razrabotki parallel'nyh programm.
URL: <http://dvm-system.org/ru/about/>
 17. *Kataev N., Kolganov A.* Additional Parallelization of Existing MPI Programs Using SAPFOR. In: *Malyshkin V. (Ed.) Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science*. 2021. Vol. 12942. Springer, Cham.
URL: https://doi.org/10.1007/978-3-030-86359-3_4
 18. *Kwon D., Han S., Kim H.* MPI backend for an automatic parallelizing compiler // *Proceedings Fourth International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN'99)*. 06.1999. P. 152–157.
 19. *Korneev V.V.* Parallel'noe programmirovaniye // *Programmnyaya inzheneriya*. 2022. T. 13, № 1. S. 3–16.
 20. *Prangishvili I.V., Vilenkin S.YA., Medvedev I.L.* Parallel'nye vychislitel'nye sistemy s obshchim upravleniem. M.: Energoatomizdat, 1983. 312 p.
 21. *Krivosheev N.M., Steinberg B.Ya.* Algorithm for searching minimum inter-node data transfers // *Procedia Computer Science, 10th International Young Scientist Conference on Computational Science, YSC 2021, 1–3 July 2021*. P. 306–313.
 22. *Gervich L.R., Steinberg B.Ya.* Automation of the Application of Data Distribution with Overlapping in Distributed Memory // *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS)*. 2023. Vol. 16, no. 1. P. 59–68.
-

23. *Shtejnberg B.Ya.* Blochno-affinnye razmeshcheniya dannyh v parallel'noj pamyati // Informacionnye tekhnologii. 2010. No 6. S. 36–41.

24. *Shtejnberg B.Ya.* Optimizatsiya razmeshcheniya dannyh v parallel'noj pamyati. Rostov-na-Donu: Izd-vo Yuzhnogo federal'nogo universiteta, 2010. 255 s.

25. *Gong Z., Chen Z., Szaday Z., Wong D., Sura Z., Watkinson N., Maleki S., Padua D., Veidenbaum A., Nicolau A.* An empirical study of the effect of source-level loop transformations on compiler stability // Proceedings of the ACM on Programming Languages. 11.2018. P. 1–29.

26. *Steinberg B.Ya., Steinberg O.B., Oganesyanyan P.A., Vasilenko A.A., Veselovskiy V.V., Zhivykh N.A.* Fast Solvers for Systems of Linear Equations with Block-Band Matrices // East Asian Journal on Applied Mathematics. 2023. Vol. 13, No. 1. P. 47–58.

27. *Vasilenko A., Veselovskiy V., Metelitsa E., Zhivykh N., Steinberg B., Steinberg O.* Precompiler for the ACELAN-COMPOS Package Solvers // In: Malyshkin V. (Ed.) Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science. 2021. Vol. 12942. P. 103–116. Springer, Cham.

URL: https://doi.org/10.1007/978-3-030-86359-3_8

28. *Dordopulo A.I., Levin I.I., Gudkov V.A., Gulenok A.A.* High-Level Synthesis of Scalable Solutions from C-Programs for Reconfigurable Computer Systems // In: Malyshkin V. (Ed.) Parallel Computing Technologies. PaCT 2021. Lecture Notes in Computer Science. 2021. Vol. 12942. Springer, Cham. URL: https://doi.org/10.1007/978-3-030-86359-3_729.

29. *Gervich L.R., Guda S.A., Dubrov D.V., Ibragimov R.A., Metelitsa E.A., Mikhailuts Y.M., Paterikin A.E., Petrenko V.V., Skapenko I.R., Steinberg B.Ya., Steinberg O.B., Yakovlev V.A., Yurushkin M.V.* How OPS (Optimizing Parallelizing System) May be Useful for Clang // CEE-SECR '2017, October 20–21, 2017, St.-Peterburg, Russian Federation. Proceedings of the 13th Central & Eastern European Software Engineering Conference in Russia ACM New York, NY, USA. 2017.

URL: <https://dl.acm.org/citation.cfm?id=3166094&picked=prox>

30. *Shtejnberg B.Ya., Shtejnberg O.B.* Preobrazovaniya programm – fundamental'naya osnova sozdaniya optimiziruyushchih rasparallelivayushchih kompilyatorov // Programmnye sistemy: teoriya i prilozheniya. 2021. T. 12, № 1(48). S. 21–113. URL: http://psta.pstiras.ru/read/psta2021_1_21-113.pdf

31. *Wolfe M.* More Iteration Space Tiling // *Supercomputing*. Reno, 1989. P. 655–664.

32. *Shtejnberg B.Ya., Shtejnberg O.B., Vasilenko A.A.* Sliyanie ciklov dlya lokalizacii dannyh // *Programmnye sistemy. Teoriya i prilozheniya*. 2020. T. 11, №3. S. 17–31. URL: <https://doi.org/10.25209/2079-3316-2020-11-3-17-31>

33. *Kozak A.V., Shtejnberg B.Ya., Shtejnberg O.B.* Algoritm vosstanovleniya sma-zannogo izobrazheniya, poluchennogo vrashchayushchejsya pod uglom k gori-zontu kameroj // *Komp'yuternaya optika*. 2020. T. 44, № 2. S. 229–235.

URL: <https://doi.org/10.18287/2412-6179-CO-598>.

URL: <http://www.computeroptics.smr.ru/>

34. *Burkhovetskiy V.V., Steinberg B.Ya.* Parallelizing an Exact Algorithm for the Traveling Salesman Problem // *Procedia Computer Science, 6-th International Young Scientist Conference on Computational Science, YSC 2017, Procedia Computer Science*. 2017. Vol. 119. P. 97-102.

URL: <http://authors.elsevier.com/sd/article/S187705091732375X>

35. *Burhoveckij V.V., Shteinberg B.Ya.* Strategiya ispol'zovaniya krupnyh zadaniy pri parallel'nom obhode dereva. Yazyki programmirovaniya i kompilyatory. Trudy Vserossijskoj nauchnoj konferencii pamyati A.L. Fuksmana. 3–5 aprelya 2017, Yuzhnyj federal'nyj universitet, g. Rostov-na-Donu: Izdatel'stvo Yuzhnogo federal'nogo universiteta, 2017. S. 66–70.

36. *Zhiyuan Li, Yonghong Song.* Automatic Tiling of Iterative Stencil Loops // *ACM Transactions on Programming Languages and Systems*. 2004. Vol. 26, No. 6. P. 975–1028.

37. *Gervich L.R., Guda S.A., Dubrov D.V., Ibragimov R.A., Metelitsa E.A., Mikhailuts Y.M., Paterikin A.E., Petrenko V.V., Skapenko I.R., Steinberg B.Ya., Steinberg O.B., Yakovlev V.A., Yurushkin M.V.* How OPS (Optimizing Parallelizing System) May be Useful for Clang // *CEE-SECR '2017, October 20–21, 2017, St.-Peterburg, Russian Federation. Proceedings of the 13th Central & Eastern European Software Engineering Conference in Russia* ACM New York, NY, USA. 2017.

URL: <https://dl.acm.org/citation.cfm?id=3166094&picked=prox>

СВЕДЕНИЯ ОБ АВТОРЕ



Штейнберг Борис Яковлевич – д. т. н, зав. каф., с. н. с. Института математики, механики и компьютерных наук Южного федерального университета.

Boris Yakovlevich STEINBERG – Doctor of Computer Science, head of chair in department of Mathematics, mechanics and computer science of Southern federal university.

e-mail: borsteinb@mail.ru

ORCID: 0000-0001-8146-0479

Материал поступил в редакцию 23 декабря 2023 года