

**ОГЛАВЛЕНИЕ**

**Л. В. Городняя**

**ФОРМЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ ПАРАДИГМАЛЬНОГО  
АНАЛИЗА ЯЗЫКОВ ПРОГРАММИРОВАНИЯ**

**А. С. Козицын, С. А. Афонин, А. А. Зензинов**

**АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПЕРЕВОДОВ СТАТЕЙ С ИСПОЛЬЗОВАНИЕМ  
СТАТИСТИЧЕСКИХ ДАННЫХ**

**А. А. Муромский, Н. П. Тучкова**

**ОБ ОНТОЛОГИИ АДРЕСАТА В МАТЕМАТИЧЕСКОЙ ПРЕДМЕТНОЙ  
ОБЛАСТИ**

**В. А. Серебряков**

**ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ВЫЧИСЛИТЕЛЬНОМ ЦЕНТРЕ  
РОССИЙСКОЙ АКАДЕМИИ НАУК – ОСНОВНЫЕ РАЗРАБОТКИ**

**С. М. Гарина, Т. В. Модянова**

**ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ЭКСПЕРТНОЙ ОЦЕНКИ ЗАЯВОК В  
КОНКУРСАХ ФЕДЕРАЛЬНОЙ ЦЕЛЕВОЙ ПРОГРАММЫ**

УДК 004.43 БК 22.18 Г

## ФОРМЫ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ ПАРАДИГМАЛЬНОГО АНАЛИЗА ЯЗЫКОВ ПРОГРАММИРОВАНИЯ

Л. В. Городняя

*Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, пр. Академика Лаврентьева, 6, Новосибирск, 630090;  
Новосибирский государственный университет, ул. Пирогова, 1, Новосибирск, 630090*

lidvas@gmail.com

### **Аннотация**

Цель статьи – выбор представления результатов сравнения языков программирования, удобного для оценки выразительной силы языков и трудоёмкости реализации систем программирования. Формы такого представления должны быть приспособлены к обоснованию практических критериев декомпозиции программ, что можно рассматривать как подход к решению проблемы факторизации весьма усложнённых определений языков программирования. Актуальность выбора лаконичных и быстро воспринимаемых форм полезен для работы в стремительно развивающемся пространстве новых проблемно-ориентированных языков программирования. Попутно можно показывать дистанцию в понятийной сложности между программированием и разработкой систем программирования.

**Ключевые слова:** *системы программирования, декомпозиция программ, реализационная прагматика, парадигмы программирования, критерии декомпозиции, семантические системы, определение языков программирования*

### **ВВЕДЕНИЕ**

Исследователи истории утверждают, что сложнее всего описывать историю человеческой мысли. В этом отношении история вычислительной техники и программирования обладает бесспорным преимуществом. Значительная часть истории мысли в этой сфере представлена в виде языков и систем программирования с корпусом программ и информационных систем, созданных на их основе, при возможности обсудить роль и путь идей с очевидцами [1]. Доступны интернет-

ресурсы, содержащие подробные описания языков программирования (ЯП), свободно распространяемые реализации систем программирования (СП), представления программ и комплектов поставки практических приложений информационных систем (ИС). Имеется ряд систематизированных интегральных представлений взаимосвязей между сотнями и тысячами ЯП, отражающих их взаимовлияние [2] и парадигмальное сходство [3]. Ценность этих обширных материалов не вызывает сомнения. Целесообразно дополнить их средствами для выражения более детальных отношений и оперативного восприятия специфики анализируемых языков и систем.

В статье рассмотрено несколько форм, нацеленных на решение проблемы лаконичного представления особенностей определения языков программирования на уровне семантических систем, примеров программ и численных характеристик. Предполагается, что, используя такие формы, можно выстроить метрическое пространство, полезное при измерении понятийной сложности конструкций, поддержанных в определениях языков и систем программирования (ЯСП). Такое пространство может быть применено при сравнении парадигм программирования (ПП), потенциала схем и моделей, используемых при разработке программ, оценки уровня новизны создаваемых ЯП, а также при выборе критериев декомпозиции программ.

### **РЕЗУЛЬТАТЫ ПАРАДИГМАЛЬНОГО АНАЛИЗА**

Парадигма программирования (ПП) как стиль мышления связана с компромиссом между особенностями решаемых задач и методами их решения с помощью программ, выполняемых автоматом. Задачи существенно различаются по следующим направлениям:

- стабильная постановка задачи, решаемой с помощью конечного автомата;
- развивающаяся постановка задачи, решаемой с помощью расширяемого автомата;
- задача бесперебойного функционирования, для решения часто требующая комплекса взаимодействующих автоматов.

Задачи каждого направления связаны противопоставлением трудно совмещаемых целей решения, таких, как академические, научно-исследовательские проблемы или производственные, практично-прикладные вопросы.

Соответствие программы и устройство конечного автомата обычно подразумевают, что имена, введённые в программе, в каждой позиции исполнения программы обладают одним значением. Для расширяемого автомата допускается возможность пополнения значений имени, включая перебор определений, представленных в программе, или их выбор по конкретному механизму. Это позволяет при развитии постановки задачи наследовать ранее отлаженный автомат.

Использование комплекса взаимодействующих автоматов позволяет компонентам декомпозированного представления программы сопоставлять отдельные автоматы, что делает возможным поддержку бесперебойного функционирования, повышения производительности программы, а также представление схем взаимодействия процессов, отвечающих трудно удостоверяемым критериям, таким как надёжность, безопасность, правильность и др. При масштабировании комплексов обычно провозглашают возможность сведения комплекса до одного автомата. Такое сведение для производственных задач может приводить к потере случаев одновременности, важных для реальной практики.

Системное программирование связано с решением задач всех таких направлений, причём одновременно и научно-исследовательской, и производственной направленности. Стабильные, желательны математические, постановки задач системного программирования обусловлены требованиями эквивалентности используемых и преобразуемых представлений программ. Развивающиеся постановки задач системного программирования связаны с динамикой совершенствования элементной базы, изменениями определений реализуемых и вновь создаваемых языков программирования, а также с ростом квалификации как разработчиков, так и пользователей программных систем. Задачи бесперебойного функционирования неизбежны для любых систем реагирования на события реального времени и массового применения, например, поддержки пользователей или организации учебного процесса.

Различия в средствах для решения разных направлений постановок задач можно выразить комплектами примеров представления программ. Анализ и сравнение большого числа ЯП разного уровня позволили выделить наиболее существенные характеристики для выражения парадигмальных различий в форме, подобной онтологическим таблицам [4]. Соответствующий пример на материале языка Lisp приведен в таблице 1.

---

Таблица 1. Парадигмальная характеристика Pure Lisp, чисто функционального подмножества языка Lisp

| <i>Параметр</i>  | <i>Описание</i>  |
|--|--|
| Эксплуатационная прагматика ЯП <sup>1</sup>                        | Язык учебного назначения, созданный специально для изучения методов функционального программирования на языке Lisp, успешно применяется при решении задач с исследовательским компонентом, требующих быстрой отладки                                   |
| Особенности системы понятий  | Все понятия сведены к разным категориям понятия «функция», унифицированы представления функций и значений, выполнение программы рассматривается как отображение списка аргументов в результат  |
| Перечень понятий, распознаваемых на уровне абстрактного синтаксиса | Символьное выражение (S-выражение), атом, вычисляемая форма, переменная, константа, ветвление, элементарные функции, определение функции, именованная функция, применение функции, аргумент функции, значение  |
| Базовые средства ЯП  | NIL, CONS, CAR, CDR, ATOM, EQ, QUOTE, EVAL, COND, LAMBDA, DEFUN  |
| Семантические расширения <sup>2</sup>                              | Работа с числами и строками рассматривается как вспомогательная семантика. Ввод и вывод данных считаются псевдо-функциями, обслуживающими отладку. Отдельное расширение языка поддерживает функции с пост-вычислением аргументов – специальные функции |
| Регистры абстрактной машины  | Стеки для хранения результатов, локальных значений переменных, вычисляемой формы и дампа, обеспечивающего защиту контекста вычислений (S E C D)  |
| Категории команд абстрактной машины                                | Засылки в стек, вычисления над стеком, пересылки из стека, ветвления, применение функции, выходы из ветвлений и функций, восстановление контекста, поддержка рекурсии  |
| Реализационная прагматика СП <sup>3</sup>                          | Списки из бинарных узлов, содержащих пару тэгированных указателей. Тэг показывает тип данного, адресуемого указателем. Автоматизировано освобождение памяти служебной программой «Сборщик мусора»  |
| Парадигмальная специфика ЯСП <sup>4</sup>                          | Исторически основополагающий классический язык, поддерживающий функциональное программирование в полном объёме   |

Аналогичный пример такой же парадигмальной характеристики языка Pascal приведён в Таблице 2.

Таблица 2. Парадигмальная характеристика подмножества языка Pascal, поддерживающего методику структурного программирования

1 Уровень языка, его ниша в полном жизненном цикле программ, соответствующая технология

2 Вспомогательные семантики, их описание относительно концептуальных языков

3 Структуры данных и традиционно подразумеваемые механизмы реализации

4 Роль языка в формировании поддерживаемой им парадигмы и/или перечень поддерживаемых парадигм

| <i>Параметр</i>  | <i>Описание</i>  |
|--|--|
| Эксплуатационная прагматика ЯП                                     | Язык учебного назначения, созданный для обучения студентов методам программирования решений задач, готовых для эффективной реализации при поддержке автоматным моделированием  |
| Особенности системы понятий  | Программа и данные – раздельные сущности. Выполнение программы сводится к шагам изменения состояний памяти, хранящей данные. Используемые в программе идентификаторы подчинены иерархии областей видимости, задаваемых вложенностью определений процедур и функций. Процедуры и типы данных не являются значениями. Возможно конструирование новых типов данных и приведение типа данных к заданному |
| Перечень понятий, распознаваемых на уровне абстрактного синтаксиса | Скаляр, вектор, значение, ключевые слова, константа, переменная, тип данных, операция, выражение, операнд, элемент вектора, сравнение, действие, последовательность действий, ветвление, цикл, определение процедуры/функции, вызов процедуры/функции  |
| Базовые средства ЯП  | TRUE, FALSE, перечислимые значения, :=, a[i ], IF, WHILE, PROCEDURE, st1 ; st2 .   |
| Семантические расширения   | Разные виды чисел, записи, множества, указатели, метки, передачи управления, функции, переключатели, конструирование типов данных, ввод-вывод, файлы, строки, библиотеки   |
| Регистры абстрактной машины  | Стеки промежуточных результатов, локальных переменных и параметров, выполняемой процедуры и вектор памяти (S E C M)  |
| Категории команд абстрактной машины                                | Засылки в стек, вычисления над стеком и памятью, пересылки из стека и локальных переменных, ветвления, передачи управления, вызовы процедур  |
| Реализационная прагматика СП                                       | Память распределяется статически по блокам заданного размера, обработка векторов использует вычисление смещений от базового адреса и подразумевает контроль границ отведенной под вектор памяти, при необходимости привлекаются библиотеки, поддерживающие ввод-вывод, работу с файлами и динамической памятью   |
| Парадигмальная специфика ЯСП                                       | Прецедент строго определения языка программирования, обладающего не слишком высокой сложностью в реализации, поддерживающего методику результативного структурного программирования. Может выполнять роль эталонного монопарадигматического языка при сравнительном анализе языков и определении их парадигматической характеристики   |

Такие таблицы могут иметь прикладное и образовательное значение, но по оперативности восприятия они заметно уступают обычным средствам деловой графики, успешно применяемым для представления сложных данных, традиционно сводимых к числовым соотношениям. Следует отметить зависимость понимания смысла от владения оттенками профессиональной терминологии и

вариациями определения понятий в разных ЯП. Современная практика программирования имеет тенденцию к использованию примеров текстов или шаблонов программ без выделения понятийной структуры. Важно, чтобы такие тексты на разных ЯП допускали ясное сопоставление и сведение к общим интуитивным понятиям. Наиболее объективные понятия такого рода связаны с архитектурными моделями, методами реализации СП и классификацией решаемых задач.

### АРХИТЕКТУРНЫЕ МОДЕЛИ

Наиболее известные компьютерные архитектуры обычно представляют как конструкцию из вычислительного устройства (E), памяти (M), устройства управления процессами (C) и средств коммуникации между устройствами и их элементами (S). Такая конструкция допускает детализацию на уровне более тонких аппаратных решений и расширение на уровне периферийных устройств, что даёт первое приближение для выбора основных видов семантических систем в языках программирования [5, 6]. Рассматривая любые семантические системы, важно отметить разницу в характере выполнения функций таких систем. Так, для любого множества значений  $V$  реализационно различимы категории функций  $F$  для методов вычислений  $FE: (V^* \rightarrow V+)^5$ , средств доступа к памяти  $FM: (T : N \rightarrow V)^6$ , особенностей управления вычислениями  $FC: (F \rightarrow \{0, 1\})^*$  и обратимой комплексации и структурирования данных  $FS: (A \rightarrow K) \cup (A \leftarrow K) = (A \leftrightarrow K)^7$ .

Интересно, что такие виды семантических систем обладают кумулятивным эффектом в порядке «VEMCS». Если  $V$  – произвольное множество значений, то  $FE: (V^* \rightarrow V+)$  – обычные вычисления, заданные формулами над значениями из этого множества, а формула может представлять самоопределимое константное значение из  $V$ . Для реализации средств доступа к памяти  $FM: (T: N \rightarrow V)$  характерно выделение понятия «адрес» ( $N$ ), и подразумевается существование специальной таблицы  $T$ , по которой определено соответствие адресов заданным значениям, при задании которых могут использоваться формулы вычислений. Особенности управления вычислениями  $FC: (F \rightarrow \{0,1\})^*$  используют разметку

---

5  $V^*$  – произвольное число значений, возможно ни одного,  $V+$  – хотя бы одно значение

6 Здесь  $N$  – множество адресов,  $T$  – таблица связей между адресами и значениями

7 Здесь  $A$  – атомарные неделимые данные,  $K$  – конструируемые структурированные данные

---

запрограммированных действий для выделения выполняемых, а также обычно понятие адресуемой памяти. Обратимая комплексация или структурирование данных в современных архитектурах FS:  $(A \leftrightarrow K)$  требуют определения границы между атомарными (A) и сложными, конструируемыми (K) объектами с возможностью как наращивания сложности, так и упрощения любых построений с учётом разных условий. Это приводит к представлению о кумулятивной шкале видов семантических систем на основе классификации архитектурно-подобных видов функций, представленных в Таблице 3.

Таблица 3. Виды семантических систем

| № столбца                  | V | E                         | M  | C  | S   |
|----------------------------|---|---------------------------|--|--|---|
| Категории функций<br>(F V) | V | FE: $V^* \rightarrow V^+$ | FM: $(T: N \rightarrow V)$<br>$V = N \cup V$ | FC: $(F \rightarrow \{0, 1\})^*$<br>$V = \{0,1\} \cup V$ | FS: $(A \rightarrow K)$<br>$U (A \leftarrow K)$<br>$V = A \cup K$ |

### ПОНЯТИЙНАЯ МАТРИЦА

Одним и тем же наборам и категориям функций могут соответствовать разные правила **R**, определяющие методы реализации СП и, следовательно, дополнительный фактор для выбора средств, иллюстрируемых с помощью текстов программ в качестве примеров отдельных понятий ЯП.

RE – обычные арифметические вычисления, отображающие произвольный ряд **значений** аргументов в не менее чем один результат,

RM – символьные вычисления, подставляющие **представления** аргументов без их предварительного вычисления,

RC – частичные или смешанные вычисления, лавирующие между вычислением и подстановкой в зависимости от разных условий,

RS – обобщённые и параллельные вычисления, оперирующие организацией процессов как множеством потоков над комплексами из разных устройств.

Представление таких различий можно выразить, дополнив горизонтальную шкалу видов функций вертикальной кумулятивной шкалой моделей вычислений или методов применения функций к значениям (R), что будет выглядеть как матрица семантических систем – понятийная матрица со строками RE, RM, RC, RS.



Строка **RE** этой матрицы представляет уровень базовой семантики ЯП (см. Таблицу 4).

Таблица 4. Семантическая декомпозиция минимального ядра ЯП

| № столбца         | <b>V</b> | <b>E</b>                 | <b>M</b>                   | <b>C</b>                        | <b>S</b>                  |
|-------------------|----------|--------------------------|----------------------------|---------------------------------|---------------------------|
| Категории функций | $V$      | $FE: V^* \rightarrow V+$ | $FM: (T: A \rightarrow V)$ | $FC: \{F \rightarrow \{0,1\}\}$ | $FS: A \leftrightarrow K$ |
| <b>RE: Ядро</b>   | Значение | Операции                 | Память                     | Управление                      | Вектор                    |

**Ядро** – семантический базис. Полное определение ЯП можно получать как консервативное расширение ядра. Обычно ядро приспособлено и к неконсервативному расширению пополнением набора библиотечных функций, реализуемых на уровне аппаратуры. Это позволяет в реализации СП для любого ЯП поддерживать разные парадигмы программирования, необходимые для поддержки полного жизненного цикла программ, чтобы достигать результата независимо от исходных возможностей ЯП.

Значение – минимальное представление объектов из области приложения языка, обычно это самоопределимые константы.

Операции – минимальный комплект функций для обработки значений.

Память – введение адресов для лаконичного и уникального представления значений (указатели, идентификаторы, переменные, метки).

Управление – разметка выполнимости композиции элементов программы из (операций, функций, действий и т. п.) специально выбранными значениями, например,  $\{0|1\}$  или  $\{\text{True} | \text{False}\}$  или  $\{\text{Nil} | T\}$ .

Вектор – обратимое конструирование одноуровневых комплектов, рассматриваемых как целостность, из которых можно восстанавливать исходные элементы. На уровне ядра достаточно одной структуры – вектора, списка, очереди и т. п.

Появление конструируемых данных является достаточным основанием для укрупнения любых видов семантических систем, что представлено строкой **RM** в таблице 5.

Таблица 5. Семантическая декомпозиция макрорасширения ядра ЯП

| № столбца         | <b>V</b> | <b>E</b>                 | <b>M</b>                   | <b>C</b>                        | <b>S</b>                  |
|-------------------|----------|--------------------------|----------------------------|---------------------------------|---------------------------|
| Категории функций | $V$      | $FE: V^* \rightarrow V+$ | $FM: (T: A \rightarrow V)$ | $FC: \{F \rightarrow \{0,1\}\}$ | $FS: A \leftrightarrow K$ |
| RE: Ядро          | Значение | Операции                 | Память                     | Управление                      | Вектор                    |
| <b>RM: Макро</b>  | Данное   | Функции                  | Задание                    | Блоки                           | Стек                      |

**Макро** – пополнение ядра средствами обработки представлений, используемых с целью укрупнения любых конструкций, что позволяет выполнять консервативное расширение ЯП. Кроме того, оно способствует лаконизму текстов программ. Макротехника позволяет наследовать отлаженность фрагментов программ. Бывает важным исключать дубли частей текста, кода и структур данных. Простейший механизм макрогенерации обычно присутствует в СП как препроцессор. Так же бывает устроена техника кодогенерации и обработки шаблонов при компиляции программ. Реализация укрупнений может быть функционально эквивалентна вызову подпрограмм. Взаимозаменяемость макроподстановки и вызова подпрограмм нередко используется при оптимизации программ.

Данное – хранимое значение или выражение, допускающее уникальность экземпляра, доступного многократно по адресу, возможно, на внешнем устройстве.

Функции – укрупнение операций с возможной параметризацией операндов. Реализационная прагматика может отличаться техникой передачи параметров через стек или специальное поле аргументов или неявно. Последнее позволяет и работу с памятью формально рассматривать как функцию с неявным аргументом, выполняющим роль функции T, задающей соответствия адресов и значений.

Задание – хранимое именованное выражение с возможностью многократного выполнения.

Блоки – хранимое выражение или код программы, представляющий составные действия, ветвления, циклы, вызовы функций, обычно с локализацией переменных, приводящей к понятию «иерархия».

Стек – схема организации данных с определённой дисциплиной доступа для поддержки иерархии, возможно с защитой независимых блоков.

Повышение уровня ЯП обеспечивается не только особым вниманием к средствам укрупнения данных на базе понятия «иерархия» и оперирование блоками программы. Дальнейшее наращивание объёмов разрабатываемых программ

отчасти достигается автоматизацией контроля некоторых условий корректности применения операций и функций к их операндам в определённых границах. Становятся важными понятия «предикат» и «тип данных», что выражено строкой RC таблицы 6. По мере расширения границ программа может стать более универсальной, способной к разумному поведению на типовых входных данных, что можно представить таблицей 6.

Таблица 6. Семантическая декомпозиция диагностического дополнения ядра ЯП

| № столбца          | <b>V</b>   | <b>E</b>                 | <b>M</b>                    | <b>C</b>                        | <b>S</b>                  |
|--------------------|------------|--------------------------|-----------------------------|---------------------------------|---------------------------|
| Категории функций  | $V$        | $FE: V^* \rightarrow V+$ | $FM: (T : A \rightarrow V)$ | $FC: \{F \rightarrow \{0,1\}\}$ | $FS: A \leftrightarrow K$ |
| RE: Ядро           | Значение   | Операции                 | Память                      | Управление                      | Вектор                    |
| RM: Макро          | Данное     | Функции                  | Задание                     | Блоки                           | Стек                      |
| <b>RC: Границы</b> | Исключения | Предикаты                | Типы данных                 | Логика                          | Варианты                  |

**Границы** – методы проверки вычислимости функций и выполнимости заданий. Цель представления границ – снижение трудоёмкости отладки программ упрощением поиска ошибок. При отсутствии ошибок проверка воспринимается как накладные расходы. Встречаются механизмы установки ловушек на непредусмотренные ситуации и программирования обработки исключений с возможностью продолжения вычислений.

Исключения – выбор специальных значений для разметки неожиданных ситуаций. В некоторых ЯП вводят значения типа “Error”. При переходе к СП происходит добавление текстовых шаблонов для формирования диагностических сообщений об исключительных ситуациях.

Предикаты – специальные функции, позволяющие определять типы значений или сравнивать значения независимо от расположения данных в памяти. Роль предиката может выполнять любая функция при подходящих договорённостях и схеме реагирования на её результаты.

Типы данных – связывание типа значения с указателем или переменной, хранящей нетипизированный код значения в памяти.

Логика – проверка соответствия типа данных или значений операциям обработки значений или доступа к памяти, возможно с учётом условий вычислимости.

Варианты – схема организации данных без определённой дисциплины доступа для организации перебора равноправных элементов или блоков. Полезно при отладке программ как механизм удостоверения принципиальной выполнимости вычислений при частичной постановке задачи.

Зависимость от учёта границ применимости программных компонент при сборке из них более сложных программ обычно преодолевается обобщением до универсальных функций, что представлено в строке RS таблицы 7. Введена дополнительная нумерация строк для более очевидного сопоставления строк и столбцов в таблицах, начиная с номера 8.

Таблица 7. Семантическая декомпозиция практического обобщения ядра ЯП

| №             | № столбцов          | V                | E                        | M                          | C                               | S                         |
|---------------|---------------------|------------------|--------------------------|----------------------------|---------------------------------|---------------------------|
| <b>строки</b> | Категории функций   | V                | FE: $V^* \rightarrow V+$ | FM: $(T: A \rightarrow V)$ | FC: $\{F \rightarrow \{0,1\}\}$ | FS: $A \leftrightarrow K$ |
| <b>E</b>      | RE: Ядро            | Значение         | Операции                 | Память                     | Управление                      | Вектор                    |
| <b>M</b>      | RM: Макро           | Данное           | Функции                  | Задание                    | Блоки                           | Стек                      |
| <b>C</b>      | RC: Границы         | Исключения       | Предикаты                | Типы данных                | Логика                          | Варианты                  |
| <b>S</b>      | <b>RS: Общность</b> | Неопределённость | Мульти-операции          | Внешний мир                | Отображения                     | Ввод-вывод                |

**Общность** – дополнительные средства обеспечения отладки и применения программ, поддерживающие возможность разумного продолжения вычислений при любых исходных данных и аварийных ситуациях.

Неопределённость – вводятся специальные дополнительные значения и ловушки (  $\_|\_$ , Error, Future). Такое расширение множества значений позволяет учитывать в текстах программ некоторые отдельные особенности процесса разработки, отладки и схемы жизненного цикла программ.

Мультиоперации – допускается произвольное число операндов операций, аргументов и результатов функций. Возможно просачивание определений на однородные структуры данных, позволяющее определения над элементарными данными автоматически распространять на более сложные данные.

Внешний мир – механизм неявного расширения области действия операций и функций на периферийные устройства, рассматриваемые как обобщение памяти.

Отображение – возможность регулярного применения функции к серии данных благодаря использованию представлений функций или указателей в качестве аргументов функций более высокого порядка.

Ввод-вывод – средства приёма данных с внешних устройств и размещения данных на внешних носителях данных, включая средства доступа к устройствам с уровня программы. Стандартно имеется в виду приём данных с клавиатуры и изображение данных на экране. Обычно подразумевается аксиоматика, требующая совместимости форматов ввода–вывода: для всякого вводимого данного существует эквивалентное ему выводимое данное и обратно – если данное может быть выведено, то его можно ввести без потерь.

Таким образом, разным категориям функций относительно схемы применения функций к аргументам на уровне ЯП при реализации СП соответствуют определённые позиции специальной понятийной матрицы, полученной как двумерная кумулятивная шкала. При анализе ЯП примеры минимальных средств его применения можно расположить по клеткам понятийной матрицы, рассматривая число или объём таких примеров как одну из численных характеристик сложности.

### **ПРИМЕРЫ ФОРМ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ СРАВНЕНИЯ ЯЗЫКОВ ПРОГРАММИРОВАНИЯ**

Результаты анализа языка Pure Lisp представлены в таблице 8, а подмножества языка Pascal – в таблице 9. В клетках таблиц расположены фрагменты текстов программ на анализируемых языках или обозначения фрагментов или понятий языка.

Понятийная матрица позволяет средства языка Pure Lisp характеризовать в рамках семантической системы, отображаемой в абстрактную машину SECD, обладающую независимыми регистрами [7]. Функции такой системы могут принадлежать основному множеству, и правило применения функций может входить в набор функций.

Таблица 8. Понятийная матрица определения языка Pure Lisp

| <i>№ стр</i> | <i>№ столбца</i> | <b>V</b> | <b>E</b> | <b>M</b> | <b>C</b> | <b>S</b> |
|--------------|------------------|----------|----------|----------|----------|----------|
|--------------|------------------|----------|----------|----------|----------|----------|

| оки      | Виды функций            | V  | F: V* → V   | AL: Atom → V   | (F → {NIL, T})*  | A ↔ K   |
|----------|-------------------------|--|---|--|--|---|
| <b>Е</b> | Ядро                    | <ul style="list-style-type: none"> <li>• NIL; и атомы</li> </ul>   | <ul style="list-style-type: none"> <li>• eq: V V → {NIL, T}</li> </ul>  | <ul style="list-style-type: none"> <li>▪ AL = (... (NIL . NIL) ( T . T ) )</li> </ul>                                    | <ul style="list-style-type: none"> <li>• Eval: Sexpr AL → V</li> <li>• quote</li> </ul>  | <ul style="list-style-type: none"> <li>▪ cons: V1 V2 → (V1 . V2)</li> <li>▪ car: (V1 . V2) → V1</li> <li>▪ cdr: (V1 . V2) → V2</li> </ul> |
| <b>М</b> | Специальные функции     | <ul style="list-style-type: none"> <li>• Sexpr: (V . V)</li> <li>• ( V . . . ) ; список</li> </ul>                                   | <ul style="list-style-type: none"> <li>• (lambda (x1 x2 ... ) form)</li> <li>• (label FN F)</li> </ul>                            | <ul style="list-style-type: none"> <li>▪ pairlis: XL YL → AL</li> <li>▪ assoc: X AL → Y</li> </ul>                       | <ul style="list-style-type: none"> <li>• ((lambda (x1 x2 ... ) form) v1 v2 ...)</li> <li>• ((label FN F) v1 v2 ...)</li> </ul> | <ul style="list-style-type: none"> <li>▪ list ; Список</li> <li>▪ (F v1 v2 . . . ) ; Форма</li> </ul>                                     |
| <b>С</b> | Границы                 | <ul style="list-style-type: none"> <li>• NIL</li> <li>• «Нет определения переменной»</li> <li>• «Нет определения функции»</li> </ul> | <ul style="list-style-type: none"> <li>• atom: V → {NIL, T}</li> <li>• apply: F V* AL → V</li> <li>• ERROR</li> </ul>             | <ul style="list-style-type: none"> <li>▪ null</li> <li>▪ Свободные переменные</li> <li>▪ Типы <b>значений</b></li> </ul> | <ul style="list-style-type: none"> <li>• cond</li> <li>• FUNCTION</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Строки</li> <li>▪ Clozure; = Funarg; = замыкание</li> </ul>                                      |
| <b>S</b> | Общность (Практичность) | <ul style="list-style-type: none"> <li>• Числа</li> <li>• Строки</li> </ul>  | <ul style="list-style-type: none"> <li>• evlis: (form* ) → V*</li> <li>• evcon: ((form form )*) → V</li> <li>• + - * /</li> </ul> | <ul style="list-style-type: none"> <li>▪ Псевдо-функции</li> <li>▪ Свойства атома</li> </ul>                             | <ul style="list-style-type: none"> <li>• map</li> <li>• reduce</li> <li>• lazy</li> </ul>                                      | <ul style="list-style-type: none"> <li>▪ READ</li> <li>▪ PRINT</li> </ul>   |

Понятийная матрица, характеризующая средства языка Pascal, обладает меньшей кумулятивностью. Pascal определён в рамках семантической системы, отображаемой в пи-код [8], в котором, в отличие от SECD, регистры SEC размещены в общей памяти. Составляющие такой системы строго разделены. Множество V фиксировано по составу, неявное определение правил организации вычислений R не допускает программируемых модификаций, функции могут выполнять роль данных в терминах указателей, связь с внешним миром определяется вне языка через библиотечные модули.

Таблица 9. Понятийная матрица определения подмножества языка Pascal

| № | № столб. | V | Е | М | С | S |
|---|----------|---|---|---|---|---|
|---|----------|---|---|---|---|---|

| строки | Виды функций | V  | F: V* → V   | M : Id → V+F   | (F → {True, False})*   | V N ↔ Array [N] of V   |
|--------|--------------|--|---|--|--|--|
| Е      | Ядро         | <ul style="list-style-type: none"> <li>• type = (a, b, ... )</li> </ul>  | <ul style="list-style-type: none"> <li>• = &lt;&gt;</li> <li>• pred succ</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Id</li> <li>▪ var X : int;</li> <li>▪ X := 1;</li> <li>▪ label L;</li> <li>▪ L: a := 1;</li> </ul>                  | <ul style="list-style-type: none"> <li>• (1*(2+X))</li> <li>• S1 ; S2</li> <li>• ключевые слова</li> <li>• goto L;</li> </ul>                    | <ul style="list-style-type: none"> <li>▪ s: array [1 .. 9] of int;</li> <li>▪ s [5] := x;</li> <li>▪ x := s [5]</li> </ul>                       |
| М      | Расширение   | <ul style="list-style-type: none"> <li>• false true</li> <li>• maxint</li> <li>• integer</li> <li>• (1,3,6,9, ...)</li> <li>• const</li> <li>• A=1;B=2;</li> <li>• type ta = 1 .. 10;</li> </ul> | <ul style="list-style-type: none"> <li>• + - * / div mod</li> <li>• F (1, x, ... )</li> </ul>   | <ul style="list-style-type: none"> <li>▪ var va : ta;</li> <li>▪ begin .</li> <li>▪ begin ... end .. end</li> <li>▪ func F subr S</li> <li>▪ end.</li> </ul> | <ul style="list-style-type: none"> <li>• if X then S1 else S2</li> <li>• begin S1 ; S2... end</li> <li>• S (X, Y)</li> <li>• forward;</li> </ul> | <ul style="list-style-type: none"> <li>▪ s: array [1 .. 9] of array [1 .. 3] of int;</li> <li>▪ s [5,2] := x;</li> <li>▪ x := s [5,2]</li> </ul> |
| С      | Границы      | <ul style="list-style-type: none"> <li>• «Нет определения»</li> <li>• “не соотв. ТД := “не соотв. ТД операнда”</li> <li>• “не соотв. ТД параметра”</li> <li>• “выход из диапазона”</li> </ul>    | <ul style="list-style-type: none"> <li>• = &lt; &gt; &gt;= &lt;=</li> <li>• F: V → {True, False}</li> <li>• in</li> <li>• not or and</li> </ul> | <ul style="list-style-type: none"> <li>▪ Типы данных:</li> <li>▪ int</li> <li>▪ char</li> <li>▪ array ...</li> </ul>   | <ul style="list-style-type: none"> <li>• case</li> <li>• Соответствие ТД</li> </ul>  | <ul style="list-style-type: none"> <li>▪ record</li> <li>▪ set</li> <li>▪ record ... case</li> </ul>   |
| С      | Практичность | <ul style="list-style-type: none"> <li>• Указатель:</li> <li>• Nil</li> <li>• X^</li> </ul>  | <ul style="list-style-type: none"> <li>• Функция-параметр</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Процедура-параметр,</li> <li>▪ /Стандартные процедуры<sup>8</sup></li> <li>▪ NEW</li> <li>▪ DESPOSE, ...</li> </ul> | <ul style="list-style-type: none"> <li>• while for until</li> </ul>  | <ul style="list-style-type: none"> <li>▪ /Стандартные процедуры<sup>9</sup></li> <li>▪ READ</li> <li>▪ WRITE</li> </ul>                          |

Такие понятийные матрицы достаточны для представления результатов сравнительного анализа ЯП и оценки сложности их применения, например, по числу различных конструкций в каждой клетке понятийной матрицы. Наполнение таких матриц можно выполнять по мере изучения отдельных ЯП. Полученные оценки можно представлять, как это сделано в Таблицах 10–12, допускающих автоматическое производство.

Таблица 10. Представление результатов сравнения понятийной сложности определений языков Pure Lisp и Pascal (число пунктов в клетках Таблиц 8, 9)

8 Библиотечное дополнение

9 Библиотечное дополнение

| Lisp |    |   |   |   |   | Pascal |    |   |      |   |      |
|------|----|---|---|---|---|--------|----|---|------|---|------|
|      | V  | E | M | C | S |        | V  | E | M    | C | S    |
| E    | 1  | 1 | 1 | 2 | 3 | E      | 1  | 2 | 5    | 5 | 3    |
| M    | 2  | 2 | 2 | 2 | 2 | M      | 7  | 2 | 5    | 4 | 3    |
| C    | 3+ | 3 | 3 | 2 | 2 | C      | 3+ | 4 | 3    | 2 | 3    |
| S    | 2  | 3 | 2 | 3 | 2 | S      | 2  | 1 | 1/3* | 1 | 0/3* |

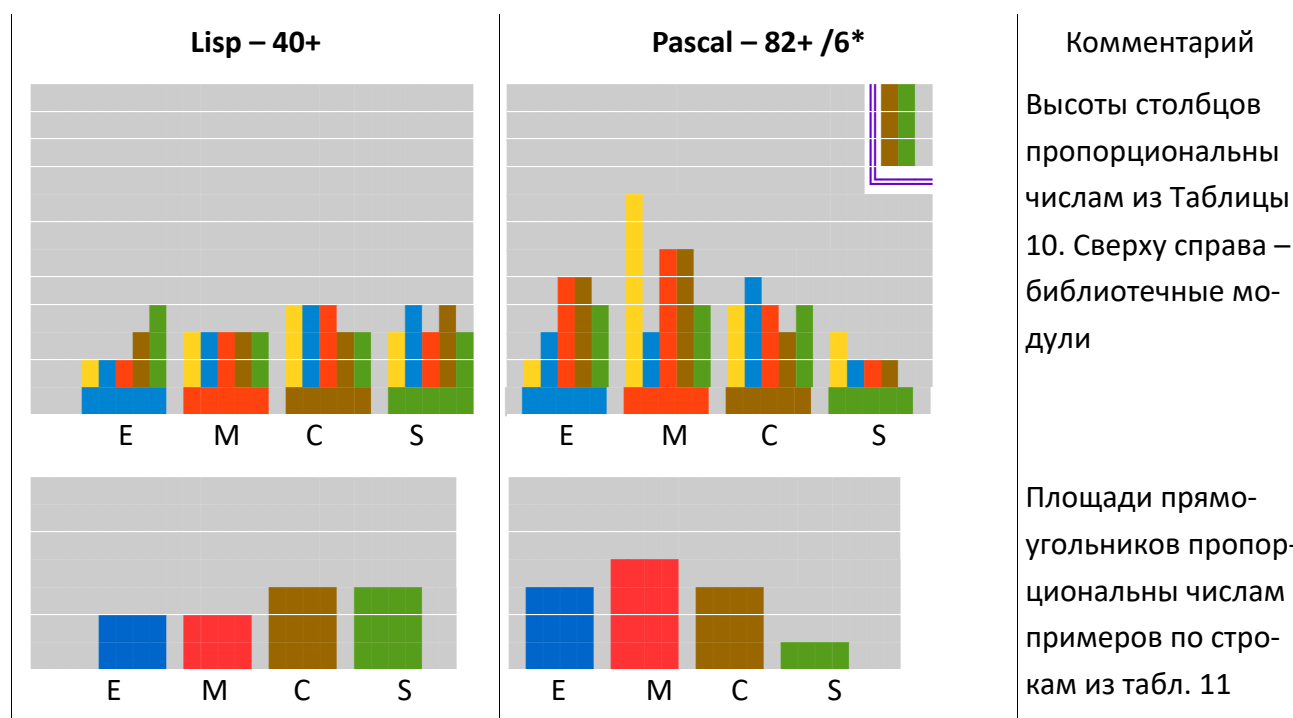
Таблица 11. Более краткие формы представления результатов сравнения сложности ЯП (число пунктов в клетках, строках и столбцах Таблиц 8, 9)

| Lisp – 43+  | Pascal – 57+ / 6*   | Комментарий   |
|---|---|---|
| ( E ( v1 e1 m1 c2 s3 )<br>M ( v2 e2 m2 c2 s2 )<br>C ( v3+ e3 m3 c2 s2 )<br>S ( v2 e3 m2 c3 s2 ) ) | ( E (v1 e2 m5 c4 s3 )<br>M (v7 e2 m5 c5 s3 )<br>C (v3+ e4 m3 c2 s3 )<br>S (v2 e1 m <b>1+3*</b> c1 s <b>0+3*</b> ) ) | Число примеров в клетках матрицы с сохранением координат клеток |
| ( E ( 1 1 1 2 3 )<br>M ( 2 2 2 2 2 )<br>C ( 3+ 3 3 2 2 )<br>S ( 2 3 2 3 2 ) )                     | ( E ( 1 2 5 4 3 )<br>M ( 7 2 5 5 3 )<br>C ( 3+ 4 3 2 3 )<br>S ( 2 1 <b>1+3*</b> 1 <b>0+3*</b> ) )                   | Число примеров в клетках матрицы с сохранением координат строк. |
| (( E 8 )<br>( M 10 )<br>( C 13+ )<br>( S 12 ) )   | (( E 15 )<br>( M 22 )<br>( C 15+ )<br>( S 5 /6* ) )   | Число примеров в строках матрицы с сохранением координат строк  |
| V E M C S<br>E 1 1 1 2 3<br>M 2 2 2 2 2<br>C 3 3 3 2 2<br>S 2 3 2 3 2                             | V E M C S<br>E 1 2 5 4 3<br>M 7 2 5 5 3<br>C 3 4 3 2 3<br>S 2 1 1 1 0   | Число примеров в клетках матрицы с сохранением координат клеток |
| Столбцы: 8 9 8 9 9  | Столбцы: 13 9 14 16 9   | Число примеров по столбцам                                      |
| Строки: 8<br>10<br>13<br>12   | Строки: 15<br>22<br>15<br>5   | Число примеров по строкам                                       |

Полученные формы позволяют ускорить сопоставление характеристик разных ЯП, с точностью до учёта соответствия координат ячеек матрицы и их наполнения. Более яркая и быстрая ассоциация получается привлечением цвета и диаграмм, подобно средствам деловой графики, используемым для представления схемно-интегральных характеристик сложных явлений.



Таблица 12. Визуальные формы представления результатов сравнения сложности ЯП



### ДИСТАНЦИЯ В ПОНЯТИЙНОЙ СЛОЖНОСТИ СЕМАНТИЧЕСКИХ СИСТЕМ

Согласно Венской методике, при определении ЯП различимы уровни абстрактной семантики, конкретного языка и конкретной машины.

1) (AC → AM) – консервативное расширение AM до AC или монотонный спуск от AC к AM, декомпозированный в соответствии с выбором вспомогательных и прикладных семантик, что делает их функционально равнозначными [9].

2) КЯ ↔ AC – отображение КЯ в AC и обратно, что делает их семантически эквивалентными.

3) AM → KM – эмуляция AM на подмножестве KM, пригодном для реализации AM, что делает их прагматически равносильными.

Это значит, что при реализации СП понятийная матрица ЯП расслаивается на три слоя, каждый из которых наследует сложность ЯП. Учитывая, что за время реализации ЯП его определение обычно претерпевает развитие, возникает необходимость системного обобщения, позволяющего сглаживать рост трудоёмкости варьирования реализации СП.

4) Системное обобщение для поддержки независимого развития разных уровней реализации СП.

E:  $\{AC \leftrightarrow AM\}$  – множество консервативных расширений AM до AC и монотонных спусков от AC до AM, декомпозированных в соответствии с выбором вспомогательных и прикладных семантик [9].

M:  $\{KY\} \leftrightarrow AC$  – множество ЯП, эквивалентных AC.

C:  $AM \leftrightarrow \{KM\}$  – множество машин, пригодных для реализации AM.

S: E+M+C+{PP} – системное обобщение можно рассматривать как концептуальное замыкание пространства конкретных ЯП, обладающих общими семантическими системами, и машин, пригодных для нетрудоёмкого переноса на них реализованных СП. Это измерение обобщает разнообразие абстрактных, изобразительных и операционных средств. Таким образом возникает прагматический реализационный куб, выглядящий как кубик Рубика с четырьмя клетками по каждому из трёх измерений (категории функций, методы вычисления, реализационные слои – уровни реализации).

Дальнейшие усложнения пространства решений, принимаемых при создании СП, направлены на инструментальную поддержку процесса разработки программ и схем, используемых при проектно-модельной подготовке этого процесса, выборе основных технологий программирования и развитии сферы приложения ЯП.

5) Технологическая поддержка (преобразования программ, факторизация, масштабирование относительно аппаратуры, вывод моделей и верификация, реорганизация и настраиваемость).

6) Человеческий фактор (индивидуальная разработка, группы, команды, фирмы-проекты).

7) Практический каркас (изученность задачи, полный жизненный цикл программ, отладка, работоспособность программы, тесты, улучшаемость, уточнение, уточнение).

8) **Эксплуатационная карта** СП – уровень пользователя (интерфейс, образцы сценариев и данных, учебные **примеры** и задачи с решениями в виде отлаженных программ, лексикон и руководства).

9) Инструментальное окружение (рабочее место системного программиста – БНФ, навигация-декомпозиция, интеграция-библиотеки компонент, систематизированных по критериям парадигмальной декомпозиции).

Можно обратить внимание, что слой 8 может содержать примеры из понятийной матрицы ЯП.

### **ЗАКЛЮЧЕНИЕ**

Описанные формы можно рассматривать как конкретизации понятийной сложности по Колмогорову [10]. Близкие работы начаты ещё в середине 1960-х годов, когда возникла задача определения ЯП, поддерживающих процессы разработки СП при создании новых ЯСП. В конце 1970-х годов в рамках работ по проекту MARC [6] был выполнен обзор динамики развития компьютерных конфигураций, что обосновывает выбор основных семантических систем [5]. В те годы был выполнен ряд серьёзных попыток создания языков системного программирования, позволяющих конструировать системы программирования [5, 11–14]. Наиболее продвинутые из них, такие, как Венская методика, не учитывали оценок трудоёмкости и понятийной сложности системного программирования, но составили базу Си-ориентированным LEX-YACC, идеи которых унаследовали Clang-LLVM.

Предложенную формализацию при оценке сложности и трудоёмкости программирования можно дополнить разделением требований к постановкам задач по сферам применения на академические и производственные, а по уровню изученности – на чёткие, развиваемые и усложнённо трудоёмкие.

### **Благодарности**

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проекты №№ 15-07-06345, 18-07-01048.

## СПИСОК ЛИТЕРАТУРЫ

1. <http://http://www.sorusom.org/> – сайт Международной конференции «Развитие вычислительной техники и ее программного обеспечения в России и странах бывшего СССР: история и перспективы».
2. <https://www.levenez.com/lang/> – сайт "Computer Languages History".
3. <http://progopedia.ru/> – сайт «Энциклопедия языков программирования».
4. *Городняя Л.В.* Парадигмы программирования: анализ и сравнение. Сиб. Отделение Рос. Акад. наук, Ин-т систем информатики им. А.П. Ершова. Новосибирск: Изд-во СО РАН, 2017. 232 с.
5. *Лавров С.С.* Методы задания семантики языков программирования // Программирование. 1978. № 6. С. 3–10.
6. *Котов В.Е.* МАРС: архитектура и языки для реализации параллелизма // Системная информатика. Вып. 1. Проблемы современного программирования. – Новосибирск: Наука. Сиб. отделение, 1991. С. 174–194.
7. *Хендерсон П.* Функциональное программирование. М.: Мир, 1983. 349 с.
8. *Вирт Н.* От Модулы к Оберону // Системная информатика. Вып 1. Проблемы современного программирования. Новосибирск: Наука. Сиб. отделение, 1991. С. 63–75.
9. *Городняя Л.В.* Резервы синтаксически ориентированного конструирования систем программирования. // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18–23 сентября 2017 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2017. С. 120–129. URL: <http://keldysh.ru/abrau/2017/proc.pdf>
10. *Колмогоров А.Н.* Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965. № 1 (1). С. 3–11.
11. *Фуксман А.Л.* Технические аспекты создания программных систем. М.: Статистика, 1979. 180 с.
12. *Koster Cornelis H.A.* Compiler Description Language. CDL3 manual. The Netherlands, August 18, 2004. <http://www.cs.ru.nl/cdl3/cdl3.pdf>
13. *Wulf W.A., Russel D.B., Habermann A.N.* BLISS: A Language for Systems Programming // CACM. 1971. V. 14. No 12. P. 780–790.
14. *Гололобов В.И., Чеблаков Б.Г., Чинин Г.Д.* Описание языка ЯРМО. Новосибирск. Препринты No 247, 248 ВЦ АН СССР, Сибирское отделение.

## FORMS OF REPRESENTATION OF RESULTS OF PARADIGMAL ANALYSIS OF PROGRAMMING LANGUAGES

L. V. Gorodnyaya

*A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Novosibirsk State University, Novosibirsk*

lidvas@gmail.com

### **Abstract**

The purpose of the article is the choice of presenting the results of a comparison of programming languages, convenient for assessing the expressive power of languages and the complexity of the implementation of programming systems. Forms of such a presentation should be adapted to substantiate practical criteria for program decomposition, which can be viewed as an approach to solving the problem of factorization of very sophisticated definitions of programming languages.

The relevance of choosing concise and quickly perceived forms is useful for working in the rapidly developing space of new problem-oriented programming languages. Along the way, you can show the distance in the conceptual complexity between programming and the development of programming systems.

**Keywords:** *programming systems, program decomposition, decomposition criteria, semantic systems, implementation pragmatics, programming paradigms, definition of programming languages*

### **REFERENCES**

1. <http://http://www.sorucum.org/> – Website of the International Conference "The development of computing equipment and its software in Russia and the countries of the former USSR: history and prospects"
2. <https://www.levenez.com/lang/> – Website of the "Computer Languages History"
3. <http://progopedia.ru/> – Website of the "Encyclopedia of programming languages"

4. *Gorodnyaya L.V.* Paradigms of programming: analysis and comparison / Rus. Acad. of Sciences, Sib. Branch, A.P. Ershov Institute of Informatics Systems. Novosibirsk: Izdat. SB RAS, 2017. 232 p.

5. *Lavrov S.S.* Methods for specifying the semantics of programming languages. // Programming. 1978. No 6. P. 3–10.

6. *Kotov V.E.* MARS: architecture and languages for realization of parallelism // System Informatics. Issue. 1. Problems of modern programming. Novosibirsk: Science. Sib. Branch, 1991. P. 174–194.

7. *Henderson P.* Functional programming. Moscow: Mir, 1983. 349 p.

8. *Wirth N.* From Modula to Oberon // System informatics. Issue 1. Problems of modern programming. Novosibirsk: Science. Sib. Separation, 1991. P. 63–75.

9. *Gorodnyaya L.V.* Reserves syntactically oriented design of programming systems // Scientific service on the Internet: works of the XIX All-Russian Scientific Conference (September 18–23, 2017, Novorossiysk). M.: IPM them. M.V. Keldysh, 2017. P. 120–129. URL: <http://keldysh.ru/abrau/2017/proc.pdf>

10. *Kolmogorov A.N.* Three approaches to the definition of "amount of information" // Problems of Information Transfer. 1965. No 1 (1). P. 3–11.

11. *Fuksman A.L.* Technical aspects of creating software systems. Moscow: Statistics, 1979. 180 p.

12. *Koster Cornelis H.A.* Compiler Description Language. CDL3 manual. The Netherlands, August 18, 2004. <http://www.cs.ru.nl/cdl3/cdl3.pdf>

13. *Wulf W.A., Russel D.B., Habermann A.N.* (1971). BLISS: A Language for Systems Programming // CACM. 1971. V. 14. No 12. P. 780–790.

14. *Gololobov V.I., Cheblakov B.G., Chinin G.D.* Description language YARMO. Novosibirsk. Preprints No 247, 248 EC of the USSR Academy of Sciences, Siberian Branch.

**СВЕДЕНИЯ ОБ АВТОРЕ**



**ГОРОДНЯЯ Лидия Васильевна** – старший научный сотрудник Института систем информатики им. А.П. Ершова СО РАН, доцент Новосибирского государственного университета, специалист в области системного программирования и образовательной информатики.

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

email: lidvas@gmail.com

*Материал поступил в редакцию 12 ноября 2018 года*

УДК 004.4

## АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПЕРЕВОДОВ СТАТЕЙ С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКИХ ДАННЫХ

А. С. Козицын<sup>1</sup>, С. А. Афонин<sup>2</sup>, А. А. Зензинов<sup>3</sup>

*НИИ механики Московского государственного университета*

*им. М.В. Ломоносова, 119991, Российская Федерация, Москва, Ленинские горы, д. 1*

<sup>1</sup>alexanderkz@mail.ru, <sup>2</sup>serg@msu.ru, <sup>3</sup>andrey.zenzinov@gmail.com

### **Аннотация**

В настоящее время происходит активное внедрение наукометрических систем для автоматизации процесса анализа эффективности деятельности научных организаций с целью применения различных методов стимулирования научной деятельности. Одними из наиболее важных индикаторов являются количество публикаций и их цитируемость. Для оценки этих показателей необходимы средства автоматизированного построения связей между оригинальными статьями и их переводами. В настоящей работе проанализированы существующие методы оценки близости оригинального текста и его возможного перевода, показана их недостаточная эффективность для построения связей между статьями и описан разработанный авторами метод автоматического поиска переводов статей в больших коллекциях библиографических данных. Особенностью разработанного алгоритма является использование статистических данных о публикации статей в различных журналах и информации о соавторах анализируемых статей. Представленный в настоящей работе алгоритм позволяет осуществлять поиск переводов статей без предварительной настройки на заданные пары языков оригинала и перевода статьи, а также не требует использования больших коллекций обучающих выборок. Апробация программной реализации алгоритма проводилась в наукометрической системе Московского государственного университета (МГУ) им. М.В. Ломоносова. Результаты тестирования показали ее достаточную эффективность и возможность использования разработанного алгоритма для автоматического построения рекомендаций пользователям для отметки в системе переводных версий статей.



**Ключевые слова:** библиографические данные, анализ графов, перевод, статья, статистика, наукометрия, цитирование, автоматизированные системы.

## **ВВЕДЕНИЕ**

В современном мире наблюдается устойчивая тенденция развития средств автоматизации оценки деятельности персонала во всех сферах экономической деятельности. Основной целью такой автоматизации является создание более эффективных механизмов поощрения и стимулирования исполнителей при решении поставленных перед ними задач. Список показателей, которые учитываются для проведения подобных оценок, существенно зависят от анализируемой области деятельности. В методике определения показателей эффективности деятельности в научной сфере (наукометрии) [1] используются многочисленные измерения и рассчитанные по ним статистические данные, учитывающие количество и качество производимой сотрудниками научной продукции: научных публикаций, докладов на конференциях, патентов и других показателей.

Учет количества подобной продукции в рамках небольшой научной организации (до 50 человек) может проводиться путем сбора бумажных отчетов о работе сотрудников или использованием простых средств автоматизации, например, таблиц в Excel.

Для больших организаций необходимо использование систем автоматизации сбора и анализа показателей научной информации. В этой области существует значительное количество небольших систем, позволяющих осуществлять централизованный сбор и предварительную обработку таких показателей [2–5], а также создаются системы корпоративного уровня, которые позволяют производить не только сбор и агрегацию информации, но и ее использование на всех уровнях управления организацией [6]. Такие системы имеют сложную архитектуру [7], включают в себя механизмы интеллектуального анализа и средства обеспечения безопасности доступа к данным [8, 9] и используют большой набор различных наукометрических показателей и индексов для оценки результатов научной деятельности.

## КОЛИЧЕСТВО ПУБЛИКАЦИЙ И ПОКАЗАТЕЛИ ЦИТИРОВАНИЯ

Традиционно одним из основных количественных показателей результатов научной деятельности являются количество публикаций и различные индексы цитирования, отражающие качество и авторитетность публикаций в научном сообществе. К основным индексам цитирования можно отнести следующие [10].

**Количество ссылок на статью.** Показатель рассчитывается по одной из баз данных Web of Science; Scopus; Google Scholar или РИНЦ. Его значение может существенно отличаться в зависимости от выбранной базы. Например, при расчете количества ссылок по РИНЦ будут учитываться ссылки из русскоязычных изданий, в то время как Web of Science больше ориентирован на англоязычные издания. Google Scholar учитывает ссылки из статей, которые размещены в интернете, но могут быть не опубликованы в журналах, и т. д. Также в различных метриках при расчете количества ссылок могут учитываться или не учитываться ссылки самоцитирования или цитирования соавторов, использоваться разные периоды времени для учета ссылок, учитываться области размещения ссылок и быть другие отличия.

**Импакт-фактор.** Показатель рассчитывается для журнала и показывает авторитетность издания, в котором опубликована статья. В отличие от количества ссылок на статью, этот показатель не оценивает востребованность самой статьи. Качество статьи оценивается в предположении, что авторитетные издания не будут публиковать статьи без серьезного рецензирования. Основным преимуществом этого показателя по сравнению с предыдущим является возможность оценки статьи непосредственно после публикации. Значение импакт-фактора рассчитывается на основе двух показателей: количества статей в журнале за предыдущие несколько лет и количества ссылок на эти статьи, сделанные в текущем году. Отношение этих величин можно назвать «средней цитируемостью» статей в журнале. Как правило, при оценках научных работ используется импакт-фактор за трехлетний период, рассчитанный по базе Web of Science.

**Индекс Хирша.** Показатель рассчитывается для автора и совмещает показатели количества статей и количества цитирований на каждую из статей. Например, значение индекса Хирша 6 означает, что у автора имеется 6 статей, которые цитируются не менее 6 раз. Значения данного показателя для одного автора также могут значительно отличаться в зависимости от базы, по которой производится расчет

количества ссылок (Web of Science, Scopus, Google Scholar или РИНЦ), учета самоцитирования и других факторов.

Для вычисления всех перечисленных выше наукометрических показателей используются число статей и число цитирований на статью. Точное определение этих параметров играет существенную роль при построении объективных наукометрических оценок. Вместе с тем, многие журналы печатают переводные версии статей, которые не является самостоятельной статьей и не должны учитываться как отдельная статья, однако ссылки, сделанные на переводную версию статьи, должны увеличивать наукометрические показатели ее авторов. В некоторых случаях такие переводы осуществляются без участия автора статьи. В этой связи возникает задача автоматизации процесса сопоставления оригинальной версии статьи и всех ее переводов.

### **ОПРЕДЕЛЕНИЕ ПЕРЕВОДОВ СТАТЕЙ**

Сложность определения переводных версий статей обусловлена тем, что предоставление информации о статье и ее переводе может осуществляться не только в разное время по мере выхода изданий, но и получаться из разных источников. Некоторые журналы на своих страницах в интернете размещают информацию о наличии у них переводных изданий, однако такая информация плохо структурирована, а ее автоматическая обработка очень сложна. Кроме того, значительная часть переводов размещается в иностранных изданиях, не являющихся полными переводными версиями соответствующих русскоязычных изданий.

Задача автоматического перевода названий статей является очень трудоемкой, поскольку в названиях используются многозначные слова, и необходимо при переводе учитывать специфику предметной области статьи. В качестве примера можно привести результаты автоматического перевода названия статьи двумя популярными переводчиками. Оригинал статьи – «Степень самоочищения агродерново-подзолистых супесчаных почв, удобренных осадком сточных вод», переводная версия статьи – «Self-Purification of Agrosoddy-Podzolic Sandy Loamy Soils Fertilized with Sewage Sludge», обратный перевод Гугл [11] – «Самоочищение Агрозодди-Подзолик Сэнди глинистые почвы, оплодотворенные с отстоем сточных вод». Оригинал статьи – «Поверхностные волны Релея и Лява при отрицательном коэффициенте Пуассона изотропных сред», переводная версия статьи – «Rayleigh and Love surface

waves in isotropic media with negative Poisson's ratio», обратный перевод Промт [12] – «Рэлей и Любовные волны поверхности в изотропических СМИ с отношением отрицательного Пуассона».

Следует отметить, что в большинстве случаев имеется смысловое сходство автоматического перевода и перевода, сделанного автором, но набор слов существенно различается. Это объясняется, в первую очередь, неоднозначностью терминов в любом языке. В одних случаях в языке перевода отсутствуют полностью эквивалентные термины языка оригинала, в других – автоматическая система выбирает не совсем верные термины.

В настоящее время, в связи усилением борьбы с плагиатом, активно развивается направление поиска эквивалентных текстов на разных языках, обсуждаемых, в том числе, на конференции «Обнаружение заимствований» [13]. Например, в системе «Антиплагиат» создан модуль «Переводные заимствования», который способен определять степень эквивалентности текстов, написанных на разных языках. Используемый в системе метод анализа основан на понятии  $n$ -грамм. Элементами  $n$ -грамм являются классы эквивалентных слов, что позволяет учитывать наличие эквивалентных терминов в разных языках [14]. Такой подход эффективен для поиска переводов полных текстов, но имеет ряд существенных недостатков, которые затрудняют его использование для поиска переводных статей по названиям.

Во-первых, построение классов эквивалентных слов требует настройки под каждую пару языков. В системе «Антиплагиат» используется только русско-английский перевод, а в случае перевода статей необходимо учитывать все возможные языки.

Во-вторых, использование  $n$ -грамм возможно только для достаточно длинных частей текста и плохо применимо к названиям статей.

Альтернативным подходом к автоматизации процесса поиска переводных версий статей является использование статистических данных о распределении статей по журналам и информации об авторах статей. Такой подход позволяет находить возможные переводы статей, основываясь только на структуре связей в графе соавторства статей, без использования информации о языке оригинала и перевода статьи. Результаты поиска могут вручную подтверждаться экспертом на основе сравнения названий статей. Использование графа соавторства успешно применяется в задачах идентификации авторов [15]. Расширение этого подхода и использование

существующих связей между статьями и журналами позволяет выявлять дополнительные закономерности в данных с целью выявления скрытых связей между объектами. Ниже представлено описание алгоритма, который определяет связи между переводными и оригинальными версиями статей на основе анализа графа соавторства.

Основой разработанного алгоритма является предположение, что оригинальная статья и ее перевод должны быть опубликованы одним и тем же авторским коллективом с разницей не более года в журналах на разных языках.

После построения пар статей, которые могут являться переводами, производится построение двудольного графа журналов, которые печатают переводные статьи. Метрика для оценки степени связи журналов в графе строится на основе мощности множеств статей в каждом из журналов и мощности множества пар статей в этих журналах, которые могут являться переводами. Результатом работы этого этапа алгоритма является множество пар журналов, в одном из которых часто печатаются переводные статьи из второго журнала. В процессе работы граф связей журналов уточняется на основе вносимых пользователями данных о своих статьях. Для этого используется как явное указание пользователями связей между оригинальной и переводной статьями, так и информация о DOI статьи, задаваемых авторами. Многие авторы указывают библиографические данные оригинала статьи в русскоязычном журнале, внося индекс DOI переводной версии для учета ссылок из Web of Science. Таким образом, собрав из внешних источников информацию о статье, по DOI можно точно определить название переводного журнала для указанного в статье русскоязычного журнала.

На основе построенного множества журналов производится поиск возможного перевода статьи. Поиск осуществляется среди статей, которые могут являться переводами (имеют совпадающее множество авторов, а дата публикации отличается не более чем на год) и опубликованы в журналах, связанных ребром в построенном ранее графе журналов.

Для апробации алгоритма использовались данные о публикациях сотрудников МГУ им. М.В. Ломоносова. Авторами статьи разработан модуль, добавленный в функционал наукометрической системы организации [16]. Разработанный для этих целей интерфейс (рис. 1) позволяет экспертам проводить оценку результатов

работы модуля и отмечать в системе правильные и ошибочные варианты предлагаемых переводов.

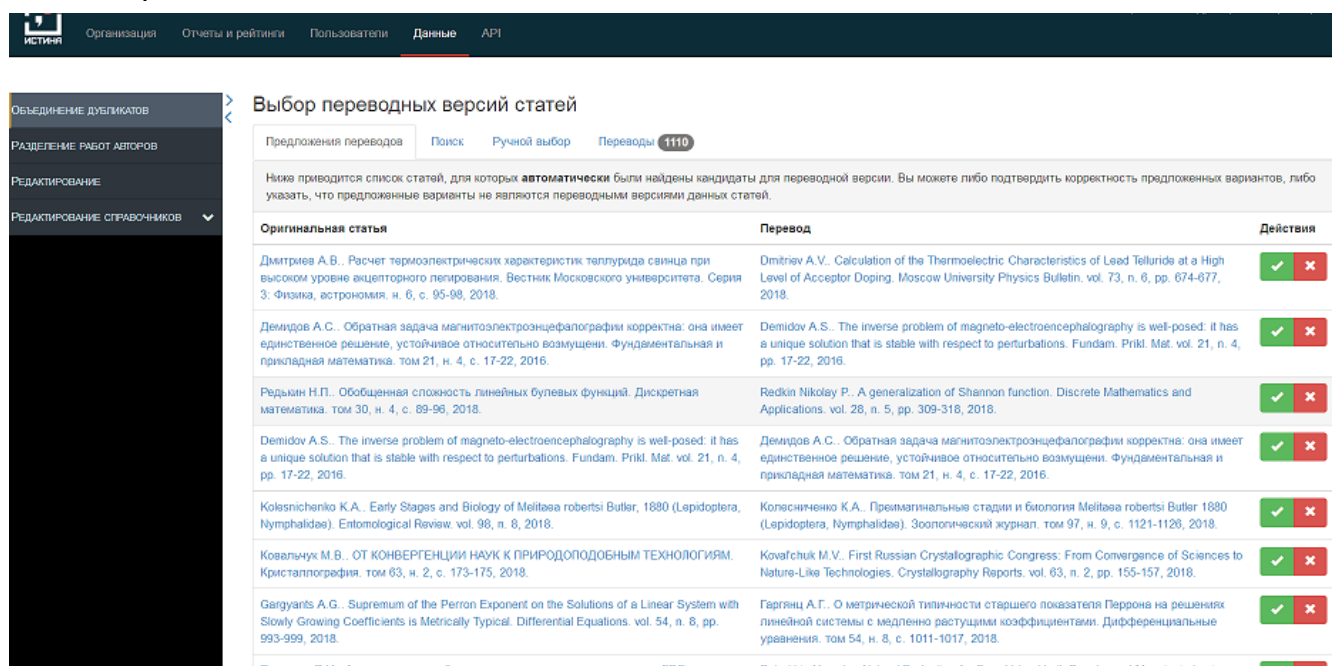


Рис. 1. Модуль выбора переводных версий статей

Для удобства работы предусмотрена возможность проведения поиска по статье, заданной пользователем (рис. 2).

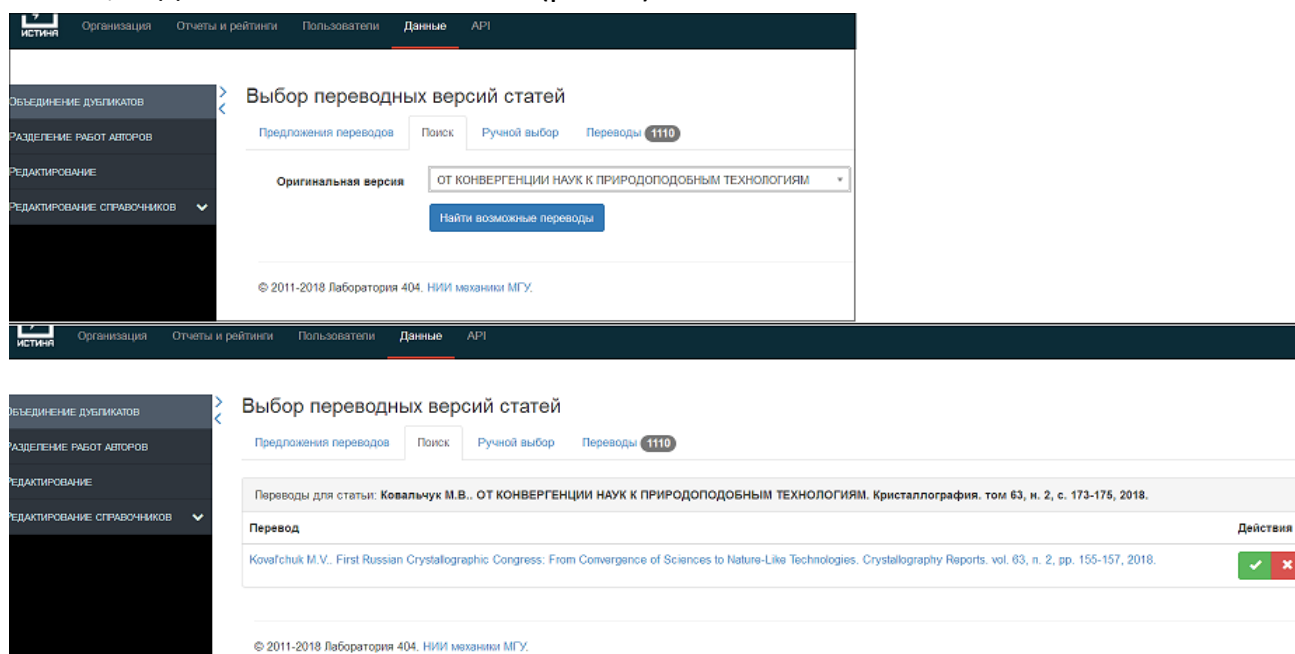


Рис. 2. Поиск по статье, заданной пользователем

Следует отметить, что алгоритм может использоваться как для обработки полной коллекции статей, так и для обработки статей, вносимых в наукометрическую информационную систему авторами непосредственно в момент их добавления. В последнем случае одним из требований является достаточная производительность реализации алгоритма, позволяющая давать рекомендации пользователю непосредственно при редактировании информации о статье в интерфейсе системы. Использование хэш-функции для множества авторов статьи позволяет производить поиск возможных вариантов перевода и давать рекомендации менее чем за 0.1 сек.

### **ЗАКЛЮЧЕНИЕ**

Разработанный нами алгоритм позволяет производить автоматический поиск переводов статей только на основе анализа графа соавторства, что обеспечивает применимость алгоритма к любым парам языков без сбора больших объемов статистической информации для построения автоматических переводов.

Результаты поиска могут использоваться самостоятельно или уточняться с использованием методов полнотекстового анализа.

Использование хэш-функции для множества соавторов позволяет значительно увеличить производительность реализации алгоритма и осуществлять поиск возможных вариантов перевода менее чем за 0.1 сек.

Результаты работы алгоритма позволяют уведомлять пользователей системы о наличии возможных переводов статей.

### **СПИСОК ЛИТЕРАТУРЫ**

1. *Налимов В.В., Мульченко З.М.* Наукометрия. Изучение науки как информационного процесса. Москва: Наука, 1969. 340 с.
2. URL: <http://www.library.spbu.ru>
3. URL: <http://library.bmstu.ru/Publications/>
4. *Алехина Е.И.* Информационная система учета научно-исследовательской деятельности сотрудников вуза // *Инновационная наука*. 2018. №5-1. С. 9–12.
5. *Столяров Р.А., Чугреев В.Л.* Автоматизированная система учета результатов интеллектуальной деятельности в научной организации. URL: <http://vtr.vscs.ac.ru/article/1512>

6. Садовничий В.А., Васенин В.А. Интеллектуальная система тематического исследования наукометрических данных: предпосылки создания и методология разработки. Часть 1 // Программная инженерия. 2018. Т. 9. №2. С. 51–58.

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

8. Васенин В.А., Иткес А.А. Внедрение реляционной модели логического разграничения доступа в web-приложения информационных систем, разработанных на основе библиотеки django // Программная инженерия. 2018. Т. 9. № 5. С. 195–208.

9. Васенин В.А., Иткес А.А., Бухонов В.Ю., Галатенко А.В. Модели логического разграничения доступа в многопользовательских системах управления наукометрическим контентом // Программная инженерия. 2016. Т. 7. № 12. С. 547–558.

10. Коряков Д.Е. Наукометрия. Зачем нужны разные индексы. URL: [https://www.mcb.nsc.ru/sites/mcb.nsc.ru/files/fck/file/naukometriya\\_2.pdf](https://www.mcb.nsc.ru/sites/mcb.nsc.ru/files/fck/file/naukometriya_2.pdf)

11. Автоматический переводчик «Гугл». URL: <https://translate.google.ru/>

12. Автоматический переводчик «Промпт». URL: <http://www.translate.ru>

13. Научная конференция «Обнаружение заимствований – 2017». URL: <http://www.oz2017.ru>

14. Плагиат в научных статьях: трудности обнаружения перевода. URL: [http://ai-news.ru/2018/01/plagiat\\_v\\_nauchnyh\\_statyah\\_trudnosti\\_obnaruzheniya\\_perevoda.html](http://ai-news.ru/2018/01/plagiat_v_nauchnyh_statyah_trudnosti_obnaruzheniya_perevoda.html)

15. Афонин С.А., Гаспарянц А.Э. Автоматическое построение функции оценки качества в задаче разрешения неоднозначности имен авторов научных публикаций // Программная инженерия. 2015. № 10. С. 31–37.

16. Наукометрическая система «ИСТИНА». URL: <https://istina.msu.ru/>

---



## ALGORITHM FOR LINKING TRANSLATED ARTICLES USING AUTHORSHIP STATISTICS

A. S. Kozitsyn<sup>1</sup>, S. A. Afonin<sup>2</sup>, A. A. Zenzinov<sup>3</sup>

*Institute of Mechanics Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991, Russian Federation*

<sup>1</sup>alexanderkz@mail.ru, <sup>2</sup>serg@msu.ru, <sup>3</sup>andrey.zenzinov@gmail.com

### **Abstract**

During the last decades scientometric techniques have been used for research activity stimulation. Number of published articles and number of their citation counts are among the most important scientometric parameters. In an automated environment, when the publications metadata is gathered from various sources, correct linking of original papers with their translations into different languages is extremely important. In the paper we show that the known text similarity measures are inefficient in the context of article linkage problem. We propose a method for semi-automatic article linkage using statistical data on authors publication activities only. This approach may be used for linking articles without training for the language of translation. The method was evaluated on real-world collection of publications metadata of ISTINA information system.

**Keywords:** *bibliographic data, graph analysis, translation, article, statistics, scientometrics, citation, automated systems.*

### **REFERENCES**

1. Nalimov V.V., Mulchenko Z.M. Naukometriia. Izuchenie nauki kak informatsionnogo protsessa. Moskva: Nauka, 1969. 340 s.
2. URL: <http://www.library.spbu.ru>
3. URL: <http://library.bmstu.ru/Publications/>
4. Alekhina E.I. Informatsionnaia sistema ucheta nauchno-issledovatel'skoi deiatel'nosti sotrudnikov vuza // Innovatsionnaia nauka. 2018. No 5-1. S. 9–12.
5. Stoliarov R.A., Chugreev V.L. Avtomatizirovannaia sistema ucheta rezultatov intellektualnoi deiatel'nosti v nauchnoi organizatsii. URL: <http://vtr.vscac.ru/article/1512>

6. *Sadovnichii V.A., Vasenin V.A.* Intellektualnaia sistema tematicheskogo issledovaniia naukometricheskikh dannyykh: predposylki sozdaniia i metodologiya razrabotki. Chast 1 // *Programmnaia inzheneriia*. 2018. V. 9. No 2. S. 51–58.

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

8. *Vasenin V.A., Itkes A.A.* Vnedrenie reliatsionnoi modeli logicheskogo razgraniicheniia dostupa v web-prilozheniia informatsionnykh sistem, razrabotannykh na osnove biblioteki django // *Programmnaia inzheneriia*. 2018. V. 9. No 5. S. 195–208.

9. *Vasenin V.A., Itkes A.A., Bukhonov V.Iu., Galatenko A.V.* Modeli logicheskogo razgraniicheniia dostupa v mnogopolzovatelskikh sistemakh upravleniia naukometricheskimi kontentom // *Programmnaia inzheneriia*. 2016. V. 7. No 12. S. 547–558.

10. *Koriakov D.E.* Naukometriia. Zachem nuzhny raznye indeksy. URL: [https://www.mcb.nsc.ru/sites/mcb.nsc.ru/files/fck/file/naukometriya\\_2.pdf](https://www.mcb.nsc.ru/sites/mcb.nsc.ru/files/fck/file/naukometriya_2.pdf)

11. Avtomaticheskii perevodchik «Gugl». URL: <https://translate.google.ru/>

12. Avtomaticheskii perevodchik «Prompt». URL: <http://www.translate.ru>

13. Nauchnaia Konferentsiia «Obnaruzhenie zaimstvovaniia – 2017». URL: <http://www.oz2017.ru>

14. Plagiat v nauchnykh statyakh: trudnosti obnaruzheniia perevoda. URL: [http://ai-news.ru/2018/01/plagiat\\_v\\_nauchnyh\\_statyah\\_trudnosti\\_obnaruzheniya\\_perevoda.html](http://ai-news.ru/2018/01/plagiat_v_nauchnyh_statyah_trudnosti_obnaruzheniya_perevoda.html)

15. *Afonin S.A., Gaspariants A.E.* Avtomaticheskoe postroenie funktsii otsenki kachestva v zadache razresheniia neodnoznachnosti imen avtorov nauchnykh publikatsii // *Programmnaia inzheneriia*. 2015. No 10. S. 31–37.

16. Naukometricheskaiia sistema «ISTINA». URL: <https://istina.msu.ru/>

## СВЕДЕНИЯ ОБ АВТОРАХ



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

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

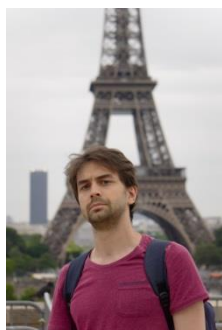
email: alexanderkz@mail.ru



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

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

email: serg@msu.ru



**ЗЕНЗИНОВ Андрей Александрович** – младший научный сотрудник, окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области моделирования распределённых информационных систем.

**Andrey Alexandrovich ZENZINOV** – Junior Researcher, graduated from M.V. Lomonosov Moscow State University. Specialist in the field of modeling of distributed information systems.

email: andrey.zenzinov@gmail.com

*Материал поступил в редакцию 26 декабря 2018 года*

УДК 519.6:004.738.52

## ОБ ОНТОЛОГИИ АДРЕСАТА В МАТЕМАТИЧЕСКОЙ ПРЕДМЕТНОЙ ОБЛАСТИ

А. А. Муромский<sup>1</sup>, Н. П. Тучкова<sup>2</sup>

*Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, 119333, ул. Вавилова, 40*

<sup>1</sup>murom@ccas.ru, <sup>2</sup>tuchkova@ccas.ru

### **Аннотация**

Обсуждена проблема представления математических предметных областей в цифровых библиотеках и полезности этих ресурсов для специалистов. Дан вариант представления математических предметных областей в интернете. В качестве информационной модели для единицы записи выбрана статья тезауруса. Реализация схемы показана на примере уравнений с частными производными. Предложен подход к организации информационного пространства автора, основанный на использовании тезауруса адресата. На основе описаний предметных областей индивидуумов предполагается построение онтологии научного междисциплинарного сообщества, что, по мнению авторов, позволит не потерять новый результат или открытие в науке, соблюсти приоритеты авторов, встроить новое знание в устоявшуюся систему классических предметных областей.

**Ключевые слова:** контролируемая лексика, дескрипторные словари, тезауруса адресата, онтология адресата

### **ВВЕДЕНИЕ**

Одна из задач научной коммуникационной среды – создание и поддержка высокого конкурентно-способного уровня научно-исследовательских работ. Если эта задача не выполняется, то в коллективе, задействованном в этой среде, нет исследований и публикаций, которые признаны актуальными в научном сообществе, по крайней мере, в некоторый временной период, или содержание информационной среды не отражает достижений в определенной предметной области, с чем чаще всего приходится сталкиваться пользователям. Однако заметим, что в

отечественном научном сообществе нет такой всеми признанной системы, как, например, [webofknowledge.com](http://webofknowledge.com) или [www.zbmath.org](http://www.zbmath.org), где собрана информация о достижениях ученых (в основном европейского и американского научного сообщества) в систематизированном виде и с учетом цитирования, что позволяет естественным образом выделить наиболее значимые и актуальные исследования, высокорейтинговые публикации и наиболее активные коллективы авторов. Наряду с общими причинами отставания российских информационных ресурсов, связанными с ориентацией научной администрации на «западные» образцы, недофинансированием проектов и пр., можно упомянуть недостаточное количество оцифрованных научных публикаций российских ученых в библиотечных системах и/или их недоступность для научного сообщества и отдельного исследователя – адресата этих знаний. Действительно, база данных, которая используется в [webofknowledge.com](http://webofknowledge.com) создавалась под руководством Ю. Гарфилда в частном Института научной информации (Institute for Scientific Information) [1] с середины 1980-х, а первые тезаурусы на основе Математических заметок Американского математического общества записывались на карточках учениками Кеннета Мея<sup>1</sup> [2], начиная с 1960-х годов. Многие знают, что Российский индекс научного цитирования (РИНЦ) на основе [e-library.ru](http://e-library.ru) начал создаваться только в начале XXI в. (при поддержке того же Ю. Гарфилда). Это свидетельствует о значительном отрыве российского сегмента цифровых научных библиотек от существующих тенденций. Отечественная фундаментальная школа в естественных науках, тем более, в математике, никогда не отставала от мирового уровня, и необходимо использовать этот багаж для представления знаний на основе цифровых технологий, разрабатывать информационные системы для специалистов отдельных предметных областей, помещать их в онтологии для интеграции накопленных знаний. В настоящей работе мы остановимся на представлении математических знаний в области задач математической физики и приведем примеры, в качестве возможных вариантов, для уравнений с частными производными.

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Kenneth\\_O.\\_May](https://en.wikipedia.org/wiki/Kenneth_O._May)

## ОНТОЛОГИЯ АДРЕСАТА

Ниже предложен подход, который базируется на создании тезауруса адресата (ТА) (термин «тезаурус адресата (индивидуума)» введен в информатику Ю.А. Шрейдером [3]) как основы для представления предметной области (ПО) автора в виде онтологии (онтологии адресата (ОА)). На особенностях тезауруса адресата в научно-информационной среде авторы подробно останавливались в работе [4]. Здесь отметим только, что в процессе информатизации могут встречаться следующие случаи:

- в результате изучения адресатом входной общественной информации ТА *изменился*, пополнился новым знанием; в этом случае можно полагать, что входная общественная информация является «информацией для адресата» (по мнению адресата);
- если же в результате изучения адресатом входной общественной информации ТА *не изменился*, то можно считать, что входная общественная информация является «неинформацией» для адресата (по мнению адресата).

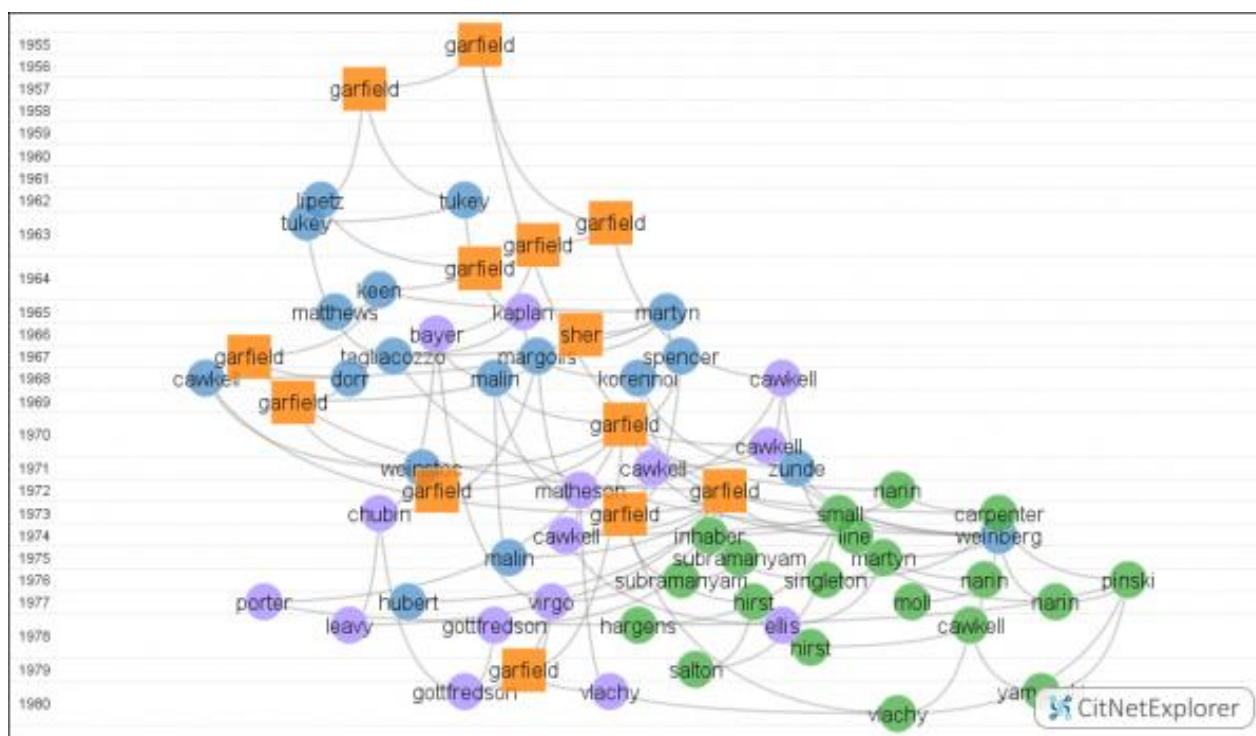


Рисунок 1. Оранжевые квадраты – работы, цитируемые Ю. Гарфилдом, остальные – работы, цитирующие Ю. Гарфилда

Например, влияние тезауруса такого адресата, как Юджин Гарфилд, показано на рис. 1, процитированном из статьи, посвященной этому выдающемуся ученому (<https://www.cwts.nl/news?article=n-q2y2c4&title=eugene-garfield-1925--2017-visionary-information-scientist>). На рисунке 1 представлена графическая иллюстрация влияния его ранних работ (1955–1980 гг.), построенная на основе WoS.

Картина влияния работ на рис. 1 отражает, в том числе, иерархические связи работ: «синие ссылки» – это цитирование работ Юджина Гарфидла, «фиолетовые» – цитирование «синих» работ, а «зеленые» – цитирование «фиолетовых». Показано влияние во втором и третьем «поколениях цитирования».

Совокупность накопленной информации в виде публикаций, отчетов, патентов и других видов научных результатов составляет множество, с помощью которого можно создать описание предметной области исследований автора или авторского коллектива.

В представлении ОА мы опираемся на множество ключевых слов (КС) из произведений автора, публикаций автора и тех публикаций, на которые он ссылается, где на него ссылаются, а также тех, которые автор включил в список своих интересов. Именно это множество составляет основу для создания ТА и модели предметной области его исследований, для краткости назовем его «информацией адресата» (с учетом того, что было сказано выше о том, что является информацией для адресата).

Например, в таблице 1 приведены русско-английские термины для дескриптора «Геллерстедта задача – The Gellerstedt problem», связанные с ним родовидовыми отношениями, выписанные на основе приведенного реферативного списка с условными идентификаторами «XXGL», где «XX» – номер, а «GL» указывает на название термина.

Это ключевые слова из цитируемых работ, объединенных заданной тематикой. С другой стороны, из реферативного списка можно получить список авторов, также объединенных данной тематикой, и включить эти ключевые слова в ТА перечисленных авторов. Также выделить «корневые» работы, основополагающие труды, где впервые встретились те и или иные термины, так как на них ссылаются все остальные авторы. При появлении поискового запроса по одному из авторов можно предлагать в позиции «см. также» работы других авторов, связанные с

этими терминами. Собственно, так устроены все современные поисковики. Особенность, на которую мы обращаем внимание, заключается в том, что выборка производится не из любого множества терминов, а из списка, ограниченного связями: реализуется механизм поиска с контролируемой лексикой из ТА, а не из истории запросов.

Таблица 1. Ключевые слова по тематике «Геллерстедта задача»

|      |   |  |
|------|---|--|
| 01GI | уравнение для Геллерстедта задачи [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]                | - equation for Gellerstedt problem [9]<br>- problem G [9]            |
| 02GI | уравнение типа Чаплыгина [8]  | - equations of Chaplygin's type [9]<br>- Chaplygin type equation [9] |
| 03GI | уравнение смешенного типа [25, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]                       | - equations of mixed type [9]  |
| 04GI | Трикоми задача [13, 15, 17, 8, 19, 22, 23]  | - Tricomi problem [9]<br>- problem T [9]                             |
| 05GI | обобщение Трикоми задачи [13, 15, 17, 8, 19, 22, 23]  | - generalized Tricomi problem [9]<br>- generalized problem T [9]     |
| 06GI | Трикоми метод [13, 15, 17, 8, 19, 22, 23]   | - Tricomi's method [9]<br>- Tricomi method [9]                       |
| 07GI | задача Коши для уравнения Трикоми [13, 15, 17, 8, 19, 22, 23]   | - Cauchy's problem or Tricomi's equation [9]                         |
| 08GI | краевая задача для уравнения типа Чаплыгина [8]   | - boundary-value problem Chaplygin type equation [9]                 |
| 09GI | краевая задача для уравнения смешанного типа [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]     | - boundary-value problem for a mixed-type equation [9]               |
| 10GI | эллиптичность уравнения для Геллерстедта задачи [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]  | - ellipticity of the equation for Gellerstedt problem [9]            |
| 11GI | параболичность для уравнения Геллерстедта задачи [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23] | - parabolicity of the equation for Gellerstedt problem [9]           |



|      |   |  |
|------|---|--|
| 12GI | гиперболичность для уравнения Геллерстедта задачи [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]                                  | - hyperbolicity of the equation for Gellerstedt problem [9]  |
| 13GI | теорема существования и единственности для уравнения Геллерстедта задачи [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]           | - existence and uniqueness theorem for the equation for Gellerstedt problem [9]  |
| 14GI | теорема существования и единственности для обобщенной задачи Трикоми [13, 15, 17, 8, 19, 22, 23]  | - existence and uniqueness theorems for the generalized Tricomi problem [9]  |
| 15GI | метод Геллерстедта [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]   | - Gellerstedt's method [9]   |
| 16GI | уравнение газовой динамики смешанного типа [8]  | - gaz dinamycal equation of mixed type [9]   |
| 17GI | приложения уравнения Геллерстедта задачи в околосзвуковой газовой динамике [8]  | - applications of the Gellerstedt problem in transsonic gas dynamics [9]   |
| 18GI | уравнение для Геллерстедта задача смешанного типа с оператором Лаврентьева-Бицадзе [5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23] | - Gellerstedt problem for the mixed type equation with Lavrentyev-Bisadze operator [9]<br>- mixed type problem G with LB operator[9] |

Например, на рис. 2 показано, что ТА автора «Моисеев Е.И.» связан с терминами из таблицы 1. В свою очередь при запросе, содержащем указанные термины, будет производиться обращение к ТА этого автора, где возможно присутствуют другие термины и связи с соавторами и дополнительная информация адресата, а КС связаны с дескрипторами предметной области, в которой работает автор.

В обобщенном виде связи из рис. 2 можно представить в виде таблицы 2 «Содержание ТА» с входами «через» автора, КС, дескрипторы ПО и публикации. Используя технологию языка онтологий, можно составить ОА.

Пополнив данные таблицы 2, можно расширить ОА. Заметим, что в виде схемы технология создания и пополнения онтологии автора выглядит достаточно просто. Однако суть проблемы представления знаний ПО и автора заключается не в перечислении публикаций и КС, а в отслеживании связей, как сделано на рис. 1. Эти связи диктуются внутренней логикой предметных областей, а также

результатами исследований, представленных в данном случае в публикациях. Именно это составляет проблемы представления ПО и математических ПО, в частности. Остановимся на одном из вариантов идентификации некоторых основных понятий ПО «математическая физика» (МФ) и смежных областей и парадигматических связей между ними.



исунок 2. Связи ТА с терминами предметной области

Таблица 2. Содержание ТА

|         |   |   |  |
|---------|---|---|--|
| ✓ Автор | <ul style="list-style-type: none"> <li>• Кс1</li> <li>• Кс2</li> <li>• Кс3</li> <li>• ...</li> <li>• КсN</li> </ul> | <ul style="list-style-type: none"> <li>➤ Термин - дескриптор1 ПО1</li> <li>➤ Термин - дескриптор1 ПО2</li> <li>➤ Термин - дескриптор1 ПО3</li> <li>➤ .....</li> <li>➤ Термин - дескрипторN ПОН</li> </ul> | <ul style="list-style-type: none"> <li>➤ Публикации</li> </ul> |
|---------|---|---|--|

### ИДЕНТИФИКАЦИЯ ОСНОВНЫХ ПОНЯТИЙ ПРЕДМЕТНОЙ ОБЛАСТИ «МАТЕМАТИЧЕСКАЯ ФИЗИКА» И СМЕЖНЫХ ОБЛАСТЕЙ

Уравнения математической физики выражают с некоторой точностью моделируемые физические или технические процессы: «Математическая модель – приближенное описание какого-либо класса явлений внешнего мира, выраженное с помощью математической символики», академик А.Н.Тихонов [24].

Учитывая основные закономерности физического процесса, мы создаем его математическую модель (там же). Например, задача о колебаниях струны, после описания определенных ограничений и допущений, приводится к приближенному уравнению

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2},$$

где  $a$  – постоянная, зависящая от физических свойств струны. Это приближение пригодно в случае, так называемых, малых колебаний струны. Уравнение носит название «волнового уравнения с двумя независимыми переменными или уравнения колебания струны».

Смежные области для ПО МФ (МФ СМобл) возникают, поскольку сама ПО существует как подкласс исследований в области теоретической физики, что делает класс рассматриваемых задач необычайно широким. Справедливо отношение:

$$\left[ \begin{array}{l} \text{Уравнения} \\ \text{математической} \\ \text{физики} \end{array} \right] \subset \left[ \begin{array}{l} \text{Методы} \\ \text{теоретической} \\ \text{физики} \end{array} \right]$$

**Физические и технические процессы.** Приведем список физических и технических процессов, которые исследуют и моделируют с помощью математических моделей МФ, опуская для краткости изложения ссылки на первоисточники – статьи Тихонова А.Н., Самарского А.А., Свешникова А.Г., Смирнова М.М., Очана Ю.С. и других классиков МФ, см. таблицу 3.

Понятие ПО МФ СМобл может быть представлено:

- *Термином* (названием);
- *Математической записью;*
- *Термином и математической записью.*

Для эффективной работы каждый термин сопровождается уникальным идентификатором с элементами мнемоники. Математическая запись в случае необходимости также сопровождается *идентификатором*.

Таблица 3. Физические и технические процессы, моделируемые в ПО МФ

|   |           |
|---|-----------|
| Задача о свободных поперечных колебаниях однородной бесконечной струны  | PrPDE0001 |
| Задача о свободных продольных колебаниях однородного стержня  | PrPDE0002 |
| Задача о свободных электрических колебаниях в проводах  | PrPDE0010 |
| Задача о колебаниях газа в бесконечной цилиндрической трубке  | PrPDE0031 |
| Задача о стационарном тепловом состоянии однородного тела   | PrPDE0040 |
| Задача о распространении тепла в ограниченном стержне, на концах которого происходит свободный теплообмен с окружающей средой                                 | PrPDE0052 |
| Задача оптимального граничного управления смещением на одном конце струны при свободном втором ее конце за произвольный достаточно большой промежуток времени | PrPDE0100 |
| Задача о деформировании изотропного линейного упругого стержня, находящегося под действием собственного веса  | PrPDE0150 |
| Задача о растяжении одноосного составного упругого тела с учетом масштабных эффектов в окрестности границы контактов  | PrPDE0160 |

Представленная выше (в таблице 3) идентификация задач позволяет оперировать записями типа «задача PrPDEXXX» или просто «PrPDEXXX» (обозначение определенной физической задачи), где используются сокращения: Pr – Problem (задача), PDE – Partial Differential Equation (уравнения с частными производными), XXXX – десятичный номер.

**Требования к терминологическим обозначениям.** В общем случае терминология предметной области обычно представлена в виде дескрипторов, недескрипторов, синонимов. Другие виды терминологии могут быть преобразованы (при необходимости) к перечисленным выше с помощью известных методов.

Приведем типичные требования и допущения при назначении / образовании терминологического обозначения понятия предметной области МФ СМобл для интернет-среды:

- словесное обозначение должно быть однозначным в рамках ПО и, вместе с тем, универсальным при использовании данного понятия;
- словесное обозначение должно быть информативным, т. е. содержать основные свойства, признаки рассматриваемого понятия, совокупность которых идентифицирует в определенной мере это понятие, но зачастую не без привнесения избыточности информации;

- словесная идентификация понятия может быть «длинной», т. е. представлять собой терминологическое словосочетание их нескольких слов-терминов;
- терминологическое обозначение может содержать специфическую информацию, имеющую определенное значение и обрамляемую круглыми скобками (сокращения, аббревиатура и др.);
- терминологическое обозначение может быть оформлено как предложение естественного языка, но без причастных оборотов, местоимений, числительных (по возможности);
- терминологическое обозначение может быть так называемым «именным понятием».

В литературе по затронутым вопросам можно отметить следующие тенденции:

а) в наименовании PDE указываются область применения уравнения, а также материальный объект физического процесса;

б) в наименовании PDE содержится в основном краткая информация о свойствах уравнения;

в) наименование PDE содержит фамилии исследователей<sup>2</sup>.

Вот характерные примеры обозначения (терминов) различных видов PDE из литературы [24] по МФ и PDE, см. пример 1 с формулами.

Пример 1.

– «уравнение свободных колебаний струны» и поясняющее «уравнение колебаний в отсутствии внешних сил»,

$$u_{tt} - \frac{1}{a^2} u_{xx} = 0, a = \sqrt{T/\rho},$$

$T$  – натяжение струны,  $\rho$  – линейная плотность струны,  $T$  и  $\rho$  – постоянные;

– «уравнение продольных колебаний однородного стержня постоянного сечения»,

$$u_{tt} - a^2 u_{xx} = 0, a = \sqrt{E/\rho_0},$$

---

<sup>2</sup> вместо словосочетания «наименование вида PDE» используем «наименование PDE»

---

$E$  – модуль упругости материала (модуль Юнга),  $\rho_0$  – объемная плотность стержня;

– «уравнение электрических колебаний в линии, свободной от искажения»,

$$v_{tt} = a^2 v_{xx}, a = 1/\sqrt{LC},$$

$L$  – коэффициент самоиндукции,  $C$  – коэффициент емкости;

– «уравнение свободных поперечных колебаний мембраны»

$$u_{tt} = a^2 (u_{xx} + u_{yy}), a = T_0/\Gamma,$$

$T_0$  – постоянная плотность натяжения во всех точках мембраны,  $\Gamma$  – постоянная поверхностная плотность мембраны;

– «уравнение колебаний»

$$\rho u_{tt} = \operatorname{div}(p \cdot \operatorname{grad} u) - q \cdot u + F(x, t),$$

неизвестная функция  $u(x, y)$  зависит от  $n(n=1,2,3)$  пространственных координат  $x=(x_1, x_2, \dots, x_n)$  и времени  $t$ ; коэффициенты  $\rho, p, q$  определяются свойствами среды, где происходит колебательный процесс;  $F(x, t)$  – свободный член, выражающий интенсивность внешнего возмущения;

– «уравнение Лаврентьева–Бицадзе»

$$u_{xx} + \operatorname{sign} y \cdot u_{yy} = 0;$$

– «волновое уравнение»

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left( \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} \right),$$

$a$  – любая постоянная (например,  $a=1$ );

– «уравнение колеблющейся струны», «уравнение плоских волн»

$$Z_{xx} - \alpha^2 Z_{tt} = 0,$$

$\alpha$  – некоторая постоянная, зависящая от массы поперечного сечения струны и ее натяжения; положив  $t = \alpha y$ , имеем

$$Z_{xx} - Z_{yy} = 0;$$

– «уравнение Трикоми», «уравнение Т»

$$yZ_{xx} + Z_{tt} = 0;$$

– «однородное волновое уравнение»

$$\left( \frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} \right) u(x, t) = 0,$$

$C$  – заданная постоянная;

– «двумерное волновое уравнение»

$$\left( \frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x_1^2} - c^2 \frac{\partial^2}{\partial x_2^2} \right) u(x_1, x_2, t) = 0,$$

$C$  – заданная постоянная.

Приведенные выше примеры – в основном PDE гиперболического типа («волновые» уравнения). Однако «характерные особенности», свойственные этим примерам, имеют место и для PDE других типов.

**Родовидовые связи некоторых задач ПО МФ.** Приведем пример 2 задач ПО МФ, где термин с идентификатором PrPDE3501 – родовой дескриптор, а PrMXPDE3502 и PrMXPDE3503 – видовые термины, «MX» – часть идентификатора, указывающая на тип уравнения, – «уравнение смешанного типа». Родовой дескриптор обозначим «Д», а видовой «→».

Пример 2.

Д: |PrPDE3501| «Задача ПО МФ»

→ |PrMXPDE3502| «Смешанная задача для волнового одномерного уравнения свободных поперечных колебаний однородной ограниченной струны с закрепленными концами»

$$\begin{aligned} u_{tt} - a^2 u_{xx} &= 0, \quad 0 < x < l, t > 0, \\ u(x, t)|_{t=0} &= u(x, 0) = \varphi(x), \quad u_t(x, t)|_{t=0} = u_t(x, 0) = \varphi(x), \\ u(x, t)|_{x=0} &= u(0, t) = 0, \quad u_x(x, t)|_{t=0} = u_{x=l}(l, t) = 0; \end{aligned}$$

→ |PrBXPDE3503| «Граничная задача для волнового одномерного

однородного уравнения с трением»

$$u_{tt} = a^2 u_{xx} - \alpha u_t \quad \alpha > 0; u(0, t) = \mu_1(t), u(l, t) = \mu_2(t).$$

В предыдущей задаче (|PMXPDE3503|) фактически для PDE мы назначили новый дескриптор: «*Волновое одномерное однородное уравнение с трением*».

Изложенный весьма краткий анализ примеров позволяет терминологически обозначить приведенную выше структуру с соблюдением определенных требований при идентификации понятий МФ и СМобл. Практически сказанное сводится к установлению родовидовых и других отношений и в массиве {PDE}, и в общем между понятиями предметной области на примере раздела «*математическое описание процессов колебаний*».

Приведем некоторые примеры, где «*Линейное неоднородное PDE второго порядка с постоянными коэффициентами и двумя независимыми переменными*» – общего характера родовой дескриптор (РД).

Пример 3.

РД: «*Линейное неоднородное PDE второго порядка с постоянными коэффициентами и двумя независимыми переменными*»

→ Одномерное однородное волновое уравнение,  $\square_{\alpha} u = 0$  (1\*).

Термин (1\*) является основанием для использования следующей терминологии:

1. «*Волновое одномерное однородное уравнение*», |PDE1046|,

$$u_{tt} = a^2 u_{xx}, u_{tt} - a^2 u_{xx} = 0, u_{xx} - \frac{1}{a^2} u_{tt} = 0, a > 0.$$

2. «*Волновое одномерное неоднородное уравнение*», |PDE1047|,

$$u_{tt} = a^2 u_{xx} + f(x, t), u_{tt} - a^2 u_{xx} = f(x, t).$$

3. «*Волновое двумерное однородное уравнение*», |PDE1048|,

$$u_{tt} = a^2 (u_{xx} + u_{yy}), u_{tt} - a^2 (u_{xx} + u_{yy}) = 0.$$

4. «*Волновое двумерное неоднородное уравнение*», |PDE1049|,

$$u_{tt} - a^2 (u_{xx} + u_{yy}) = f(x, y, t), u_{tt} - a^2 (u_{x_1 x_1} + u_{x_2 x_2}) = f(x_1, x_2, t).$$

5. «*Волновое n-мерное неоднородное/однородное уравнение*»,

---



$$u_{tt} - a^2 \sum_{i=1}^n u_{x_i x_i} = \Phi(x_1, x_2, \dots, x_n, t) / \Phi = 0.$$

В приложениях обычно  $1 < n < 3, a > 0$  – постоянная.

#### 6. «PDE гиперболического типа»

Все эти термины могут рассматриваться как родовые дескрипторы для перечисленных выше в примере 3 (причем для п. 6 только именно так).

Для термина «волновое одномерное однородное уравнение» как родового дескриптора имеем видовые дескрипторы, см. пример 4.

Пример 4.

Д: «*волновое одномерное однородное уравнение*»

→ «*волновое одномерное однородное уравнение свободных поперечных колебаний бесконечной/ограниченной однородной струны*» (для ограниченной струны, например, «с закрепленными концами»);

→ «*волновое одномерное однородное уравнение свободных продольных колебаний однородного ограниченного стержня постоянного сечения с закрепленными концами*»;

→ «*волновое одномерное однородное уравнение электрических колебаний в линии без искажений*».

«*Волновое n-мерное неоднородное/однородное уравнение*» – видовые дескрипторы по отношению к дескриптору «*PDE гиперболического типа*», см. пример 5.

Пример 5.

Д: «*PDE гиперболического типа*»

→ «*Волновое n-мерное неоднородное/однородное уравнение*»

Приведенные дескрипторы (см. пп. 1–4 примера 3) являются видовыми по отношению к дескриптору п. 5 соответственно при  $\Phi \neq 0$  и  $\Phi = 0$ .

**Состав статьи тезауруса «математической задаче ПО МФ», основные сведения.** Приведем информацию о математической задаче ПО МФ в таблице 4 «основных сведений». Более подробно об особенностях в связи с описанием парадигматических отношений ПО МФ сказано в работе [25].

Таблица 4. Основные сведения математической задаче ПО МФ

|   |
|---|
| <p><u>Терминологическое обозначение</u> – название данной задачи<br/>                 дескриптор данной задачи: [&lt;Название задачи&gt;&lt;Дескриптор PDE&gt;<br/>                 &lt;Информация о физическом процессе&gt; &lt;Сведения о материальной объекте&gt;]</p>   |
| <p>Термины-синонимы</p>   |
| <p>Литература</p>   |
| <p><u>Математическая запись задачи</u><br/>                 Дескриптор PDE. Идентификатор PDE с указанием типа PDE<br/>                 Запись PDE в декартовой системе координат<br/>                 Запись PDE в используемой системе координат (отличной от декартовой)<br/>                 Начальные условия<br/>                 Граничные условия. Однородные, неоднородные условия<br/>                 Интерпретация независимых переменных</p> |
| <p><u>Корректность постановки задачи</u><br/>                 Утверждение о существовании и единственности решения задачи<br/>                 Непрерывная зависимость решения задачи от исходных данных (условий)</p>  |
| <p><u>Решение задачи</u><br/>                 Точное решение задачи в виде формульной зависимости (аналитическое решение).<br/>                 Название точного решения<br/>                 Приближенное решение задачи (численное решение). Название приближенного решения.<br/>                 Табличное решение задачи</p>  |
| <p><u>Методы решения задачи</u><br/>                 Методы точного решения<br/>                 Методы приближенного решения<br/>                 Методы табличного решения</p>  |
| <p><u>Связи</u><br/>                 &lt;физические   технические задачи&gt; → &lt;задача предметной области МФ&gt; :<br/>                 идентификаторы, ссылки</p>   |
| <p>Ключевые слова:</p>  |
| <p>См. также:</p>   |
| <p>Историческая справка:</p>  |
| <p>Примечание: поясняющая и дополняющая информация</p>  |

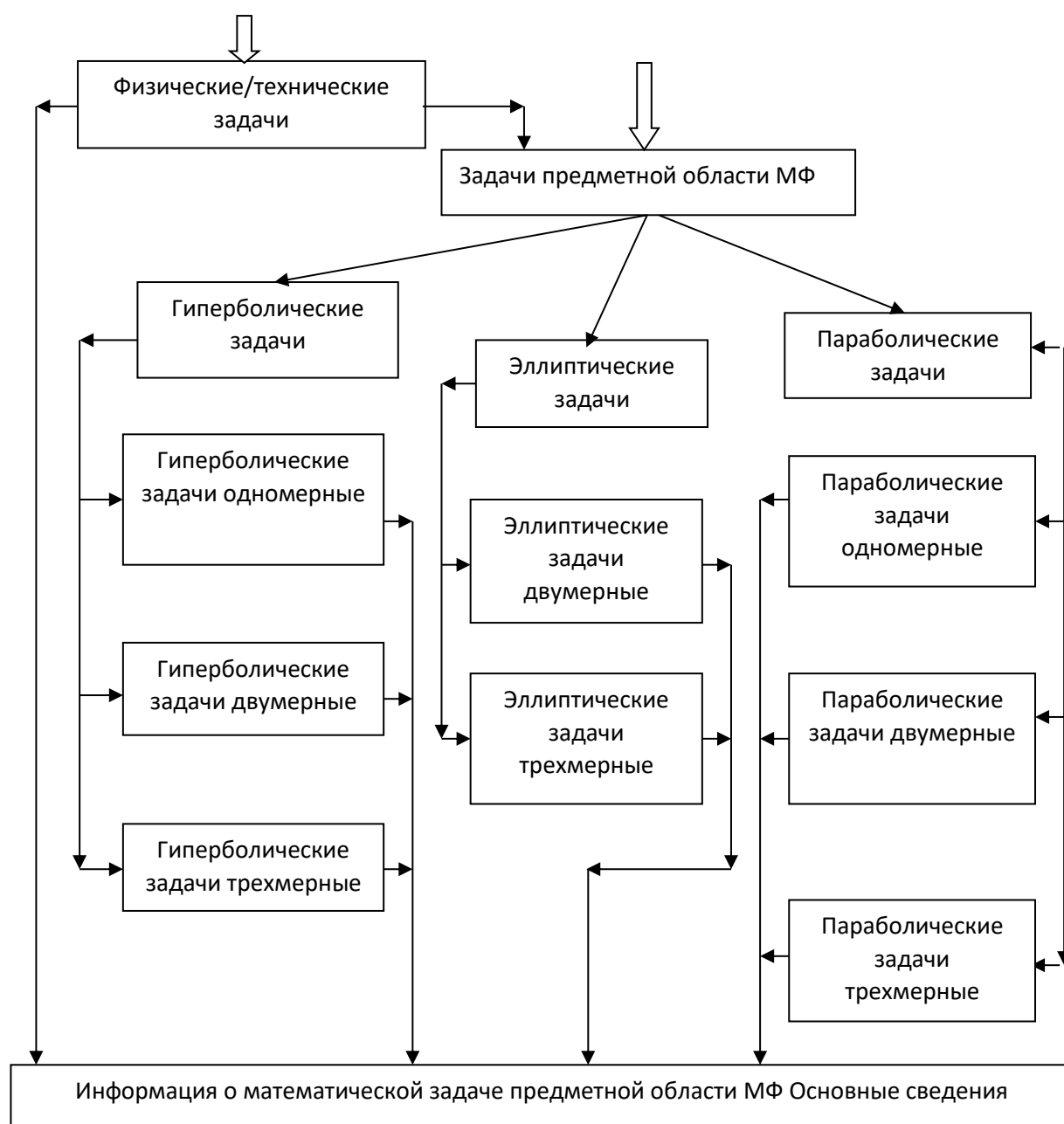


Рисунок 3. Схема иерархических связей задач МФ

В «Основных сведениях» могут быть предусмотрены различные входы в информационный массив конкретной задачи МФ. Тем не менее, нелишне иметь общую иерархическую систему связей задач МФ. На рис. 3 приведен один из возможных вариантов, где «...мерность» определяется пространственными переменными.

Приведем также схематически родовидовые отношения для PDE второго порядка с двумя независимыми переменными, см. рис. 4.

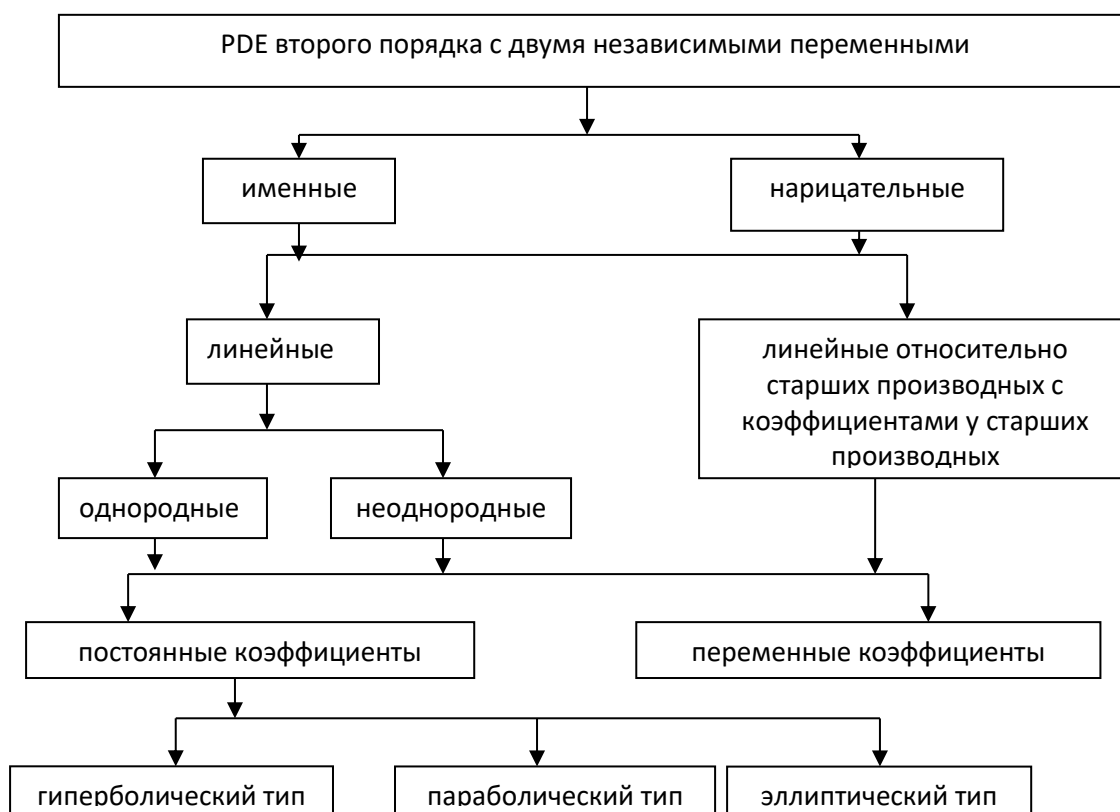


Рисунок 4. Схема связей PDE второго порядка с двумя независимыми переменными

**Дескрипторы некоторых разделов ПО МФ.** На основании схемы рис. 3 можно выписать дескрипторы раздела «Задачи ПО МФ», см. таблицу 5.

Таблица 5. Дескрипторы раздела «Задачи ПО МФ»

|                |  |
|----------------|--|
| ✓ Задачи ПО МФ | PH1PDE4500  «Гиперболические задачи одномерные»<br> PH2PDE4499  «Гиперболические задачи двумерные»<br> PH3PDE4498  «Гиперболические задачи трехмерные»<br> PP1PDE4495  «Параболические задачи одномерны»<br> PP2PDE4494  «Параболические задачи двумерные»<br> PP3PDE4493  «Параболические задачи трехмерные»<br> PE2PDE4497  «Эллиптические задачи двумерные»<br> PE3PDE4496  «Эллиптические задачи трехмерные» |
|----------------|--|

Многие физические процессы подразделяются на этапы, для каждого из которых может быть использована модель, воплощаемая задачей из приведенного списка. Далее приведем дескрипторы с математическими записями для

некоторых конкретных задач для родового дескриптора «Задачи смешанного типа», см. таблицу 6.

Таблица 6. Дескрипторы раздела «Задачи смешанного типа»

|                                 |  |
|---------------------------------|--|
| <p>✓ Задачи смешанного типа</p> | <ul style="list-style-type: none"> <li>•  PMXPDE3506  «Смешанная задача для одномерного однородного уравнения теплопроводности о распределении температуры в ограниченном, однородном, с тепловой изоляцией, на концах которого задан тепловой режим»             <ul style="list-style-type: none"> <li>○ <math>u_t - a^2 u_{xx} = 0, \quad 0 &lt; x &lt; l, t &gt; 0;</math></li> <li>○ <math>u(x, 0) = \varphi(x), 0 \leq x \leq l,</math></li> <li>○ <math>u(0, t) = \alpha(t), u(l, t) = \beta(t), t \geq 0.</math></li> </ul> </li> <li>•  PMXPDE3505  «Смешанная задача для волнового одномерного неоднородного уравнения вынужденных поперечных колебаний ограниченной однородной струны, закрепленной на концах»             <ul style="list-style-type: none"> <li>○ <math>u_{tt} = a^2 u_{xx} + f(x, t), \quad 0 &lt; x &lt; l, t &gt; 0;</math></li> <li>○ <math>u(x, 0) = \varphi(x), u_t(x, 0) = \psi(x),</math></li> <li>○ <math>u(0, t) = 0, u(l, t) = 0.</math></li> </ul> </li> <li>•  PMXPDE3507  «Смешанная задача для волнового двумерного однородного уравнения свободных поперечных колебаний однородной прямоугольной мембраны с закрепленными краями»             <ul style="list-style-type: none"> <li>○ <math>u_{tt} = a^2 (u_{xx} + u_{yy}), \quad 0 &lt; x &lt; l, 0 &lt; y &lt; m, t &gt; 0;</math></li> <li>○ <math>u(x, y, t) _{t=0} = \varphi(x, y), \quad u_t(x, y, t) _{t=0} = \psi(x, y),</math></li> <li>○ <math>u(0, y, t) = 0, u(l, y, t) = 0,</math></li> <li>○ <math>u(x, 0, t) = 0, u(x, m, t) = 0.</math></li> </ul> </li> <li>•  PMXPDE3508(pol)  «Смешанная задача для волнового однородного уравнения свободных поперечных колебаний однородной круглой мембраны с закрепленной границей»             <ul style="list-style-type: none"> <li>○ <math>u_{tt} = a^2 (u_{xx} + u_{yy});</math></li> <li>○ <math>u_{tt} = \frac{a^2}{r} \left( \frac{\partial u}{\partial r} + r \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial^2 u}{\partial \varphi^2} \right), \quad 0 \leq r \leq l, 0 &lt; \varphi &lt; 2\pi, t &gt; 0;</math></li> <li>○ <math>u(l, \varphi, t) = 0,</math></li> <li>○ <math>u(r, \varphi, 0) = f(r, \varphi),</math></li> <li>○ <math>u_t(r, \varphi, 0) = \psi(r, \varphi).</math></li> </ul> </li> </ul> |
|---------------------------------|--|

**Статья тезауруса с заглавным дескриптором PDE.** По результатам анализа приведенных примеров приведем вариант содержания статьи тезауруса с заглавным дескриптором PDE для общего случая, см. таблицу 7.

Таблица 7. Состав статьи тезауруса с заглавным дескриптором данного PDE (общий случай)

|   |
|---|
| Терминологическое обозначение – дескриптор данного PDE  |
| Термины - синонимы  |
| Литература  |
| Запись данного PDE в декартовой системе координат (с указанием геометрической (пространственной) и иной интерпретации независимых переменных) |
| Структура данного PDE в декартовой системе координат: главная часть; младшие члены; правая часть.   |
| Запись данного PDE в полярной системе координат $(r, \varphi)$ с указанием ссылки об использовании  |
| Запись данного PDE в цилиндрической системе координат $(r, \varphi, z)$ с указанием ссылки об использовании                                   |
| Запись данного PDE в сферической системе координат $(\rho, \theta, \varphi)$ с указанием ссылки об использовании                              |
| Запись данного PDE в криволинейной системе координат $(q_1, q_2, q_3)$ с указанием ссылки об использовании                                    |
| Запись данного PDE с помощью понятий и символики векторного анализа с указанием ссылки об использовании                                       |
| Характеристическое уравнение. Характеристики PDE  |
| Каноническая форма данного PDE с указанием канонического преобразования   |
| Тип данного PDE: HT, ET, PT, MX   |
| Общее решение данного PDE с указанием метода получения этого решения, а также ссылок  |
| Нетривиальное частное решение данного PDE с указанием ссылок. При необходимости ограничиваются только ссылками.                               |
| Идентификаторы математических задач, в которых данное PDE используется (в любой записи)   |
| Родовые дескрипторы (по крайней мере, ближайший)<br>Видовые дескрипторы (по крайней мере, ближайший)<br>Ассоциативные дескрипторы             |
| Ключевые слова  |
| См. также   |
| Историческая справка  |
| <u>Примечание:</u> пояснения, ссылки, связи, дополнительная необходимая информация  |

В таблице 7 обозначено: HT – уравнения гиперболического, ET – эллиптического, PT – параболического, MX – смешанного типов.

Характерные особенности приведенных примеров в определенной мере типичны для рассматриваемой предметной области. Представленные здесь «общее и частное» – констатация на информационном уровне, она позволяет перейти к терминологическому рассмотрению PDE второго порядка с двумя независимыми переменными с дальнейшей детализацией при указании типов уравнений.

**PDE второго порядка с двумя независимыми переменными.** Имеем следующие термины:

- «Линейное неоднородное/однородное PDE второго порядка с постоянными коэффициентами и двумя независимыми переменными»

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu = \begin{cases} f(x, y), & | \text{PDE1007} | \\ 0, & | \text{PDE1003} | \end{cases}, \quad (1)$$

где  $A, B, C, D, E, F$  – постоянные;

- «Линейное неоднородное/однородное PDE второго порядка с переменными коэффициентами и двумя независимыми переменными»

$$a_{11}(x, y)u_{xx} + 2a_{12}(x, y)u_{xy} + a_{22}(x, y)u_{yy} + b_1(x, y)u_x + b_2(x, y)u_y + b_3(x, y)u = \begin{cases} f(x, y), & | \text{PDE1023} | \\ 0, & | \text{PDE1019} | \end{cases}, \quad (2)$$

$a_{11}, a_{12}, a_{22}, b_1, b_2, b_3$  – функции только переменных  $x, y$  и с определенной степенью гладкости;

Дескриптор «Линейное PDE второго порядка с двумя независимыми переменными» |PDE1001|, математическая запись уравнения состоит из выражений (1), (2). Этот дескриптор (с идентификатором |PDE1001|) является родовым по отношению к каждому из дескрипторов, уравнения которых представлены в выражениях (1) и (2).

- «Линейное относительно старших производных PDE второго порядка с постоянными коэффициентами у старших производных (с правой частью/без правой части) и двумя независимыми переменными»

$$A_1u_{xx} + 2B_1u_{xy} + C_1u_{yy} + F_1(x, y, u, u_x, u_y) = \begin{cases} \varphi(x, y), & | \text{PDE1015} | \\ 0, & | \text{PDE1011} | \end{cases}, \quad (3)$$

где  $A_1, B_1, C_1$  – постоянные;

- «Линейное относительно старших производных PDE второго порядка с переменными коэффициентами у старших производных (с правой частью/без правой части) и двумя независимыми переменными»

$$A_2(x, y)u_{xx} + 2B_2(x, y)u_{xy} + C_2(x, y)u_{yy} + F_2(x, y, u, u_x, u_y) = \begin{cases} \varphi_1(x, y), & | \text{PDE1031} | \\ 0, & | \text{PDE1027} | \end{cases} \quad (4)$$

где  $A_2, B_2, C_2$  – достаточно гладкие функции только переменных  $x, y$ ;

Дескриптор «Линейное относительно старших производных PDE второго порядка с двумя независимыми переменными» | PDE1002 |, математическая запись уравнения содержит PDE из выражений (3) и (4).

Дескриптор с идентификатором | PDE1002 | является родовым по отношению к каждому из дескрипторов, уравнения которых представлены в выражениях (3) и (4).

Функции  $F_1$  и  $F_2$  – неглавные части уравнений естественно рассматривать (как один из вариантов) таким образом, чтобы выражения (3) и (4) не дублировали линейные выражения (1) и (2).

Далее: «Уравнение в частных производных называется квазилинейным, если оно линейно относительно всех старших производных искомой функции».

Термин «Квазилинейное дифференциальное уравнение второго порядка с двумя независимыми переменными» используется для PDE вида

$$a \frac{\partial^2 u}{\partial x^2} + 2b \frac{\partial^2 u}{\partial x \partial y} + c \frac{\partial^2 u}{\partial y^2} + \Phi(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}) = 0,$$

где коэффициенты  $a, b, c$  – функции переменных  $x, y$  и принадлежат классу  $C^2$  в некоторой области переменных.

В общем случае имеем дескрипторы:

– «Квазилинейное PDE второго порядка с двумя независимыми переменными (с правой частью/без правой части)»

$$A_3(x, y, u, u_x, u_y)u_{xx} + 2B_3(x, y, u, u_x, u_y)u_{xy} + C_3(x, y, u, u_x, u_y)u_{yy} + F_3(x, y, u, u_x, u_y) = \begin{cases} \varphi_2(x, y), & | \text{PDE1044} | \\ 0, & | \text{PDE1045} | \end{cases} \quad (5)$$

Список дескрипторов далеко не полон и отражает взгляды авторов на проблему представления тезауруса ПО МФ СМобл в цифровой среде. В дальнейших



исследованиях предполагается продолжить эти работы по идентификации терминов, включая методы решений и решения задач этой ПО.

### **ЗАКЛЮЧЕНИЕ**

Предложен один из подходов к решению проблемы поиска в больших информационных ресурсах, основанный на представлении предметных областей с помощью их терминов, которые в дальнейшем используются в поисковых запросах. При успешной реализации этот подход обеспечивает уменьшение информационного шума, времени обработки запроса и повышает эффективность поиска публикаций из заданной предметной области. Однако остается открытым вопрос о последующем информационном терминологическом наполнении моделей предметных областей, что связано с естественным развитием науки и техники. Использование информационного пространства пользователя, автора, работающего в некоторой предметной области, дает возможность пополнять знания предметной области через тезаурус адресата. Рассмотрен вариант идентификации понятий математической предметной области, задач и дескрипторов предметной области уравнений с частными производными, состав статьи тезауруса с заглавным дескриптором из этой предметной области, который реализуется на базе цифровой библиотеки LibMeta [26].

### **Благодарности**

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проекты № 17-07-00217 и № 18-07-00297.

### **СПИСОК ЛИТЕРАТУРЫ**

1. *Cawkell T., Garfield E.* Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar H. Fredriksson (Ed.). IOS Press, 2001. P. 149–160.
2. *Lewis A.C.* Kenneth O. May and Information Retrieval in Mathematics // *Historia Mathematica*. 2004. No 31 (2). P. 186–195.
3. *Мусеев Е.И., Муромский А.А., Тучкова Н.П.* Онтология научного пространства или как найти гения // *Онтология проектирования*. 2014. №4 (14). С. 18–33.
4. *Шрейдер Ю.А.* Тезаурусы в информатике и теоретической семантике // *Научно-техническая информация*. Сер. 2. 1971. № 3. С. 21–24.

5. *Gellerstedt S.* Doctoral Thesis, 1935; Jbuch Fortschritte Math. 61, 1259.
6. *Rassias J.M.* Lecture Notes on Mixed Type Partial Differential Equations. World Scientific, 1990, 144 p.
7. *Трикоми Ф.Д.* Лекции по уравнениям в частных производных, пер. с итал., М.: Изд-во иностранной литературы, 1957. 446 с.
8. *Смирнов М.М.* Уравнения смешанного типа. М., 1970. 296 с.
9. *Smirnov M.M.* Equations of mixed type. American Translation of the Mathematical monographs. Vol. 51. Mathematical Soc., 31 Dec 1978. P. 232.
10. *Моисеев Е.И., Таранов Н.О.* Решение одной задачи Геллерстедта для уравнения Лаврентьева–Бицадзе // ДУ. 2009. Т. 45, № 4. С. 543–548.
11. *Моисеев Е.И., Таранов Н.О.* Интегральное представление решения одной задачи Геллерстедта // ДУ. 2009. Т. 45, № 11. С. 1554–1559.
12. *Моисеев Е.И., Лихоманенко Т.Н.* Об одной нелокальной краевой задаче для уравнения Лаврентьева–Бицадзе // ДАН. 2012. Т. 446, № 3. С. 256–258.
13. *Moiseev E.I., Nefedov P.V.* Tricomi problem for the Lavrent'ev–Bitsadze equation in a 3d domain // IT&SF. 2012. Vol. 23, No 10. P. 761–768.
14. *Moiseev E.I., Nefedov P.V.* Gellerstedt problem for the Lavrent'ev–Bitsadze equation in a 3D-domain // IT&SF. 2014. Vol. 25, Issue 7. P. 509–512.
15. *Моисеев Е.И., Холومهева А.А., Нефедов П.В.* Аналоги задач Трикоми и Франкля в трехмерных областях для уравнения Лаврентьева–Бицадзе // ДУ. 2014. Т.50, № 12. С. 1672–1675.
16. *Moiseev E.I., Nefedov P.V., Kholomeeva A.A.* Analog of the Gellerstedt problem for the Lavrent'ev–Bitsadze equation in a 3D domain Differential Equations // Differential Equations. 2015. Vol. 51. No. 6. P. 827–829.
17. *Moiseev E.I., Moiseev T.E., Vafodorova G.O.* On an Integral Representation of Neumann–Tricomi Problem for the Lavrent'ev–Bitsadze Equation // Differential Equations. 2015, Vol. 51. No. 8. P. 1065–1071.
18. *Моисеев Е.И., Лихоманенко Т.Н.* Собственные функции задачи Трикоми с наклонной линией изменения типа // ДУ. 2016. Т. 52, № 10. С. 1375–1382.
19. *Zarubin A.N., Kholomeeva A.A.* Tricomi problem for an advance-delay equation of mixed type with variable deviation of the argument // Differential Equations. 2016. Vol. 52, No. 10. P. 1312–1322.
20. *Моисеев Е.И., Моисеев Т.Е., Холومهева А.А.* О разрешимости задачи

Геллерстедта с данными на параллельными характеристиках // ДУ. 2017. Т. 53, N 10. С. 1379–1384.

21. *Moiseev E.I., Likhomanenko T.N.* Eigenfunctions of the gellerstedt problem with an inclined-type change line // IT&SF. 2017. Vol. 28, No 4. P. 328–335.

22. *Moiseev E.I., Likhomanenko T.N.* Eigenfunctions of the tricomi problem with an inclined type change line // Differential Equations. 2016. Vol. 52, No 10. P. 1323–1330.

23. *Moiseev E.I., Gulyaev D.A.* The completeness of the eigenfunctions of the Tricomi problem for the Lavrent'ev–Bitsadze equation with the Frankl gluing condition // IT&SF. 2016. Vol. 27. No 11. P. 893–898.

24. *Тихонов А.Н., Самарский А.А.* Уравнения математической физики. Изд-е 2-е, исправл. и дополн. М.: ГЛАВЛИТ, 1953. 679 с.

25. *Моисеев Е.И., Муромский А.А., Тучкова Н.П.* Интернет и математические знания: представление уравнений математической физики в информационно-поисковой среде. М: Изд-во МАКС Пресс, 2008. 80 с.

26. *Серебряков В.А., Атаева О.М.* Информационная модель открытой персональной семантической библиотеки LibMeta // Труды XVIII Всероссийской научной конференции «Научный сервис в сети Интернет». Новороссийск, 19–24 сентября 2016 г. ИПМ им. М.В. Келдыша РАН. С. 304–313.

---

## ABOUT ONTOLOGY OF THE ADDRESSEE IN MATHEMATICAL SUBJECT DOMAIN

A. A. Muromskiy<sup>1</sup>, N. P. Tuchkova<sup>2</sup>

*Dorodnicyn Computing Centre of Federal Research Centre "Computer Science and Control" of Russian Academy of Sciences, Moscow 119333. Vavilov str., 40*

<sup>1</sup>murom@ccas.ru, <sup>2</sup>tuchkova@ccas.ru

### **Abstract**

The problem of representation of mathematical subject domains in digital libraries and usefulness of these resources for experts is discussed. The option of representation of mathematical subject domains on the Internet is given. As information model for unit of record article of the thesaurus is chosen. Implementation of the scheme is shown on the example of the partial differential equations. Approach to the organization of information space of the author is offered, using the thesaurus by the addressee. On the basis of descriptions of subject domains of individuals creation of ontology of scientific cross-disciplinary community is supposed what, according to authors, will allow not to lose new result or opening in science, to observe priorities of authors, to build in new knowledge the settled system of classical subject domains.

**Keywords:** *controlled lexicon, descriptor dictionaries, thesaurus by the addressee, ontology of the addressee*

### **REFERENCES**

1. *Cawkell T., Garfield E.* Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar H. Fredriksson (Ed.). IOS Press, 2001. P. 149–160.
2. *Lewis A.C. Kenneth O.* May and Information Retrieval in Mathematics // *Historia Mathematica*. 2004. No 31 (2). P. 186–195.
3. *Moiseev E.I., Muromskiy A.A., Tuchkova N.P.* Ontologia nauchnogo prostranstva ili kak nayti geniya // *Ontologiya proektirovaniya*. 2014. No 4 (14). S. 18–33.
4. *Shrejder Ju.A.* Tezaurusy v informatike i teoreticheskoy semantike // *Nauchno-tehnicheskaja informacija*. Ser. 2, 1971. No 3. S. 21–24.
5. *Gellerstedt S.* Doctoral Thesis, 1935; *Jbuch Fortschritte Math.* 61, 1259.

6. *Rassias J.M.* Lecture Notes on Mixed Type Partial Differential Equations. World Scientific, 1990. 144 p.

7. *Tricomi F.D.* Lekcii po uravnenijam v chastnyh proizvodnyh, per. s ital., M.: Izd-vo inostrannoj literatury, 1957, 446 s.

8. *Smirnov M.M.* Uravnenija smeshannogo tipa. M., 1970. 296 s.

9. *Smirnov M.M.* Equations of mixed type. American Translation of the Mathematical monographs. Vol. 51. Mathematical Soc., 31 Dec 1978. P. 232.

10. *Moiseev E.I., Taranov N.O.* Reshenie odnoj zadachi Gellerstedta dlja uravnenija Lavrent'eva–Bicadze // Differential Equations 2009. Vol. 45, No 4. S. 543–548.

11. *Moiseev E.I., Taranov N.O.* Integral'noe predstavlenie reshenija odnoj zadachi Gellerstedta // Differential Equations 2009. Vol. 45. No 11. S. 1554–1559.

12. *Moiseev E.I., Lihomanenko T.N.* Ob odnoj nelokal'noj kraevoj zadache dlja uravnenija Lavrent'eva–Bicadze // DAN. 2012. Vol. 446. No 3. S. 256–258.

13. *Moiseev E.I., Nefedov P.V.* Tricomi problem for the Lavrent'ev–Bitsadze equation in a 3d domain // IT&SF. 2012. Vol. 23, No 10. S. 761–768.

14. *Moiseev E.I., Nefedov P.V.* Gellerstedt problem for the Lavrent'ev–Bitsadze equation in a 3D-domain // IT&SF. 2014. Vol. 25, Issue 7. S. 509–512.

15. *Moiseev E.I., Holomeeva A.A., Nefedov P.V.* Analogi zadach Triкоми i Franklja v trehmernyh oblastjah dlja uravnenija Lavrent'eva–Bicadze // Differential Equations 2014. Vol. 50, No 12. S. 1672–1675.

16. *Moiseev E.I., Nefedov P.V., Kholomeeva A.A.* Analog of the Gellerstedt problem for the Lavrent'ev–Bitsadze equation in a 3D domain Differential Equations // Differential Equations. 2015. Vol. 51. No 6. S. 827–829.

17. *Moiseev E.I., Moiseev T.E., Vafodorova G.O.* On an Integral Representation of Neumann–Tricomi Problem for the Lavrent'ev–Bitsadze Equation // Differential Equations 2015. Vol. 51. No 8. S. 1065–1071.

18. *Moiseev E.I., Lihomanenko T.N.* Sobstvennyye funkicii zadachi Triкоми s naklonnoj liniej izmenenija tipa // Differential Equations 2016. Vol. 52. No 10. S. 1375–1382.

19. *Zarubin A.N., Kholomeeva A.A.* Tricomi problem for an advance-delay equation of mixed type with variable deviation of the argument // Differential Equations. 2016. Vol. 52, No 10. S. 1312–1322.

20. *Moiseev E.I., Moiseev T.E., Holomeeva A.A.* O Razreshimosti zadachi

Gellerstedta s dannymi na parallel'nymi harakteristikah // *Differential Equations*. 2017. Vol. 53, No 10. S. 1379–1384.

21. *Moiseev E.I., Likhomanenko T.N.* Eigenfunctions of the gellerstedt problem with an inclined-type change line // *IT&SF*. 2017. Vol. 28. No 4. S. 328–335.

22. *Moiseev E.I., Likhomanenko T.N.* Eigenfunctions of the tricomi problem with an inclined type change line // *Differential Equations*. 2016. Vol. 52. No 10. S. 1323-1330.

23. *Moiseev E.I., Gulyaev D.A.* The completeness of the eigenfunctions of the Tricomi problem for the Lavrent'ev–Bitsadze equation with the Frankl gluing condition // *IT&SF*. 2016. Vol. 27. No 11. S. 893–898.

24. *Tihonov A.N., Samarskij A.A.* *Uravnenija matematicheskoy fiziki*. Izd-ie 2 e, ispravl. i dopoln. M.: GLAVLIT, 1953. 679 s.

25. *Moiseev E.I., Muromskij A.A., Tuchkova N.P.* *Internet i matematicheskie znanija: predstavlenie uravnenij matematicheskoy fiziki v informacionno-poiskovoj srede*. M: Izd-vo MAKS Press, 2008. 80 s.

26. *Serebrjakov V.A., Ataeva O.M.* *Informacionnaja model' otkrytoj personal'noj semanticheskoy biblioteki LibMeta* // *Trudy XVIII Vserossijskoj nauchnoj konferencii "Nauchnyj servis v seti internet"*. Novorossijsk, 19–24 Sep. 2016. IPM im. M.V. Keldysha RAN. S. 304–313.

## СВЕДЕНИЯ ОБ АВТОРАХ



**МУРОМСКИЙ Александр Александрович**, ст. н. с., ВЦ РАН, к. ф.-м. н., окончил мехмат МГУ им. М.В. Ломоносова и университет им. Н.Э. Баумана, долгие годы работал в ВИНТИ. Специалист в области математического анализа и информационных технологий.

**Alexander Alexandrovich MUROMSKIY**, senior researcher of CCAS, PhD., graduated from mechanics and mathematics faculty of Lomonosov MSU and the university of N.E. Bauman, for many years worked in VINITI. The expert in the field of the mathematical analysis and information technologies.

e-mail: murom@ccas.ru



**ТУЧКОВА Наталья Павловна**, ст. н. с. ВЦ РАН, к. ф.-м. н., окончила ВМиК МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

**Natalia Pavlovna TUCHKOVA**, senior researcher of CCAS, PhD., graduated from CS faculty of Lomonosov MSU. The expert in the field of algorithmic languages and information technologies.

e-mail: tuchkova@ccas.ru

*Материал поступил в редакцию 10 декабря 2018 года*

УДК 004

## ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ВЫЧИСЛИТЕЛЬНОМ ЦЕНТРЕ РОССИЙСКОЙ АКАДЕМИИ НАУК – ОСНОВНЫЕ РАЗРАБОТКИ

В. А. Серебряков<sup>1</sup>

*Вычислительный центр им. А.А. Дородницына Федерального  
исследовательского центра «Информатика и управление» Российской  
академии наук, г. Москва, 119333, ул. Вавилова, 40*

<sup>1</sup> serebr@ultimeta.ru

### **Аннотация**

Рассмотрены основные проекты, которые были реализованы в Вычислительном центре им. А.А. Дородницына Российской академии наук (ВЦ РАН) за последние 20 лет, т. е. с 1998 года. Одним из первых был реализован пилотный проект «Интегрированной системы информационных ресурсов (ИСИР) РАН». Успешное завершение этого проекта позволило развернуть работы по интеграции разнородных научных информационных ресурсов в общеакадемическую научную информационную систему. Важным этапом был проект создания Единого Научного Информационного Пространства (ЕНИП) РАН. Этот проект основывался на подсистеме «Научный институт РАН», созданной в ВЦ РАН и Центре научных телекоммуникаций (ЦНТК) РАН. Учитывая важность формирования цифровых библиотек, Российская академия наук приняла в 2006 году целевую научную программу «Создание ЦБ «Научное наследие России»», в соответствии с которой была реализована цифровая библиотека. Созданный портал «ГеоМета» – это стандартизированная и децентрализованная среда управления пространственной информацией, разработанная для доступа к базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников, облегчающая обмен пространственной информацией между организациями и ее совместное использование посредством интернета.

В настоящее время основное направление работ – цифровая персональная семантическая библиотека LibMeta. Основная задача этой системы заключается в предоставлении пользователю унифицированного представления для



возможности автоматизированного извлечения интересующей его информации по определенной предметной области.

**Ключевые слова:** предметная область, научная предметная область, научная информация, научные знания, обобщенное представление научной предметной области, таксономии, тезаурусы, глобальные онтологии, поисковые системы, организация научных знаний, цифровые библиотеки

## **ВВЕДЕНИЕ**

Работы в направлении создания систем, интегрирующих информационные ресурсы Российской академии наук (РАН), были начаты в 1998 году. Благодаря поддержке Межведомственной программы «Национальная сеть компьютерных телекоммуникаций для науки и высшей школы» был реализован пилотный проект «Интегрированная система информационных ресурсов (ИСИР) РАН». Успешное завершение этого проекта позволило развернуть работы по интеграции разнородных научных информационных ресурсов в общеакадемическую научную информационную систему. В 2001 году по инициативе Отделения математики РАН была принята новая программа целевых расходов Президиума РАН «Информатизация научных учреждений и Президиума РАН». Главной задачей этой программы стала поэтапная интеграция информационных ресурсов организаций РАН в объединенное информационное пространство – Единую информационную систему (ЕИС) РАН. Координация этих работ осуществлялась Советом РАН «Научные телекоммуникации и информационная инфраструктура». Основная часть работ по собственно разработке системы была выполнена в Отделе систем математического обеспечения Вычислительного центра (ВЦ) РАН и Отделе информационных технологий Центра научных телекоммуникаций (ЦНТК) РАН. Первоочередной задачей проекта ЕИС РАН стала разработка концептуальной основы и инфраструктуры для интеграции разнородных информационных и вычислительных ресурсов организаций РАН в единое информационное пространство. Единое информационное пространство (информационную инфраструктуру фундаментальных и прикладных исследований РАН) должны составлять всевозможные цифровые библиотеки, информационные и вычислительные системы организаций РАН, использующие как собственные принципы организации, так и, по возможности, технологию открытой архитектуры проекта ЕИС или непосредственно ее релизы. В

результате был подготовлен системный проект, который определил структуру системы, как таковой, типы информационных ресурсов, участвующих в системе, общую функциональность компонентов системы. В проекте также были отражены принципы организации распределенной системы и интеграции в систему уже существующих ресурсов.

Важным этапом был проект создания Единого Научного Информационного Пространства (ЕНИП) РАН. Этот проект основывался на подсистеме «Научный институт РАН», созданной в ВЦ РАН и ЦНТК РАН. Эта подсистема обеспечивает возможность интеграции информационных ресурсов отдельных организаций в ЕИС. На базе этой системы были реализованы веб-информационные системы ряда институтов и отделений РАН, а также такие информационные системы, как Научное наследие России, портал интеграции пространственных данных Геомета и ряд других. Система была запущена в опытную эксплуатацию.

### **1. ИНТЕГРИРОВАННАЯ СИСТЕМА ИНФОРМАЦИОННЫХ РЕСУРСОВ (ИСИР) РАН**

С 2001 года выполнялась целевая программа Президиума РАН «Информатизация научных учреждений и Президиума РАН» (с 2004 года – «Информатизация»). С самого начала официальной деятельности по программе значительные усилия были приложены к выработке согласованного системного взгляда на стоящие проблемы и пути их решения, к формированию целей и задач, подходов к решению, базовых требований к используемым методам (технологиям, стандартам и т. п.). В связи с этим с участием всех заинтересованных сторон был разработан ряд документов, положенных в основу большинства проектов, выполняющихся в рамках Программы. В целом надо сказать, что до определенного момента все процессы, связанные с применением информационных технологий в РАН, двигались полностью бессистемно, не управлялись и не контролировались. Тем более не было никакого анализа полученных результатов, эффективности вложений и т. п. С момента деятельности рабочих групп по программе информатизации эта бессистемность постепенно начала исчезать. Кроме бессистемности, проблемы были еще и такими:

- отсутствие полного понимания, согласованного со всеми заинтересованными сторонами в РАН, необходимости развития работ в направлении интеграции;

- как следствие, задержки при окончательной формулировке и принятии общей концепции и программы работ по информатизации РАН;
- отсутствие юридической базы, которая могла бы создать условия для защиты авторских прав и прав интеллектуальной собственности на разработки, выполняемые в РАН;
- различные уровни подготовленности организаций РАН к внедрению и использованию современных ИТ;
- отсутствие или недостаточная подготовленность к интеграции базовых информационных блоков, которыми должны быть информационные системы Институты, Центральные библиотек, Отделений и Президиума РАН;
- отсутствие адаптированных к требованиям РАН разработок в области стандартизации объектов и механизмов единой системы;
- как следствие, отсутствие возможности полноценного обмена информацией в электронном виде.

Общая задача ИСИР РАН [1, 2] состоит в организации единого информационного пространства. Это требует решения задач по извлечению и структуризации метаданных, по обеспечению их ввода в структурированном виде. Второй класс задач состоит в предоставлении средств интеграции информации разнообразных информационных систем (репозиториев), тем или иным способом накопивших структурированную информацию. Модель данных представлена на рис. 1. С точки зрения пользователя ИСИР представлена как Портал РАН. Портал реализован (совместно с ИПИ РАН) как Информационно-поисковый справочник РАН, ориентированный на накопление и предоставление оперативной научно-административной информации. В настоящий момент основными типами ресурсов справочника (Портала РАН) являются следующие:

- Организации РАН в соответствии со структурным делением РАН (президиум, отделения, секции, научные центры, филиалы РАН) и сведения о них;
  - Сотрудники РАН (аппарат РАН, аппарат отделений, руководство организаций и учреждений, научные сотрудники) и сведения о них (адреса, телефоны и т.д.);
  - Публикации;
  - Проекты.
-

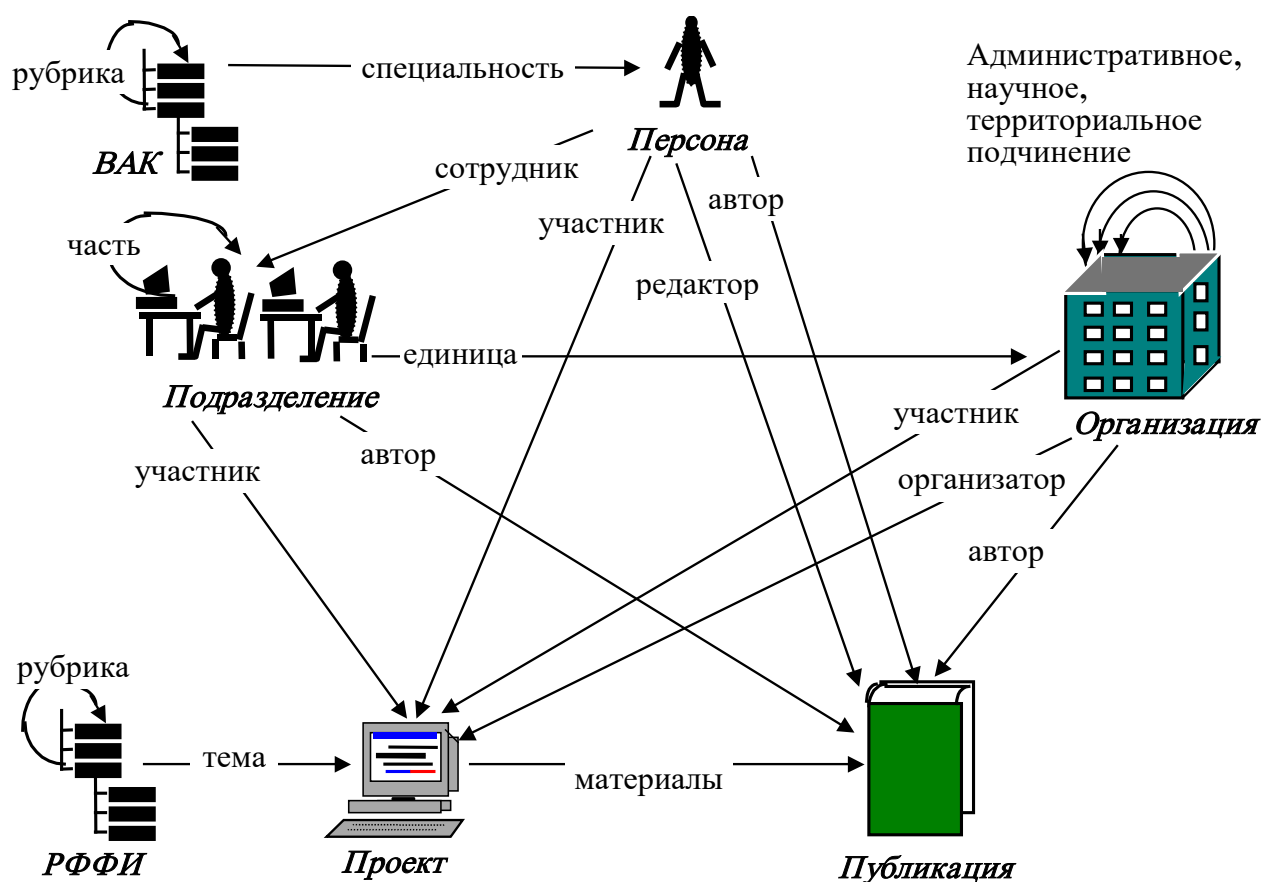


Рисунок 1. Модель данных ИСИР РАН

Справочник РАН отражает организационно-структурное деление РАН, позволяет получить информацию о структурных подразделениях РАН и обеспечивает доступ к информационным ресурсам этих подразделений, данным о сотрудниках учреждения, их научной деятельности. Исполнительная система справочника обеспечивает следующие возможности:

- Просмотр информации и средства навигации по структуре информации;
- Поиск информации по различным видам запросов и просмотр выданной по запросам информации;
- Средства ввода, редактирования и сопровождения информации;
- Средства администрирования непосредственно в подразделениях РАН.

При выборе платформы для был произведен анализ существующих инструментальных средств, пригодных для создания подобной системы. Выбор был сделан в пользу платформы ASP.NET, как обеспечивающей максимальную производительность, удобные средства разработки, компонентную ориентированность, открытость и расширяемость архитектуры, позволяющую вмешиваться практически во все этапы обработки поступающих веб-запросов. Текущее состояние портала представлено на рис. 2.



Рисунок 2. Портал РАН

## 2. ИНФОРМАЦИОННАЯ СИСТЕМА НАУЧНЫЙ ИНСТИТУТ (НИ) РАН

Рассмотрим типовой научный институт, входящий в состав РАН. Он представляет собой полноценную организацию со сложной административной структурой, основным направлением деятельности являются научные исследования.

Задачи, решаемые каждой такой структурной единицей РАН, можно разделить по своему типу – административные, научные, публичные и т. д.

**Административные задачи.** В любой организации для нормального функционирования требуется постоянное решение управленческих задач, влияющих прямым образом на деятельность организации в целом и выполнение конкретных задач на всех уровнях. Сюда входят такие задачи, как управление организационной структурой и кадрами, управление проектами, обеспечение документооборота и пр.

**Научные задачи.** Основным направлением деятельности любого научного института РАН являются научные исследования, а основной задачей организации становится в этой плоскости обеспечение научной деятельности сотрудников.

**Публичные задачи.** Взаимодействие с другими научными учреждениями, организация и проведение конференции и научных семинаров, публикация научных трудов сотрудников, предоставление доступа к результатам научных экспериментов, научным данным – все это составляет неотъемлемую часть деятельности научного института. Информационная система Института РАН должна, с одной стороны, стать центром научно-информационного сервиса сотрудников Института, а с другой, – обеспечивать полное представление информации о научной деятельности Института для мирового сообщества. Информационная система Института РАН должна представлять собой узел в распределенной архитектуре множества узлов – информационных систем Институтов РАН. На основе описанных выше задач научных организаций в составе РАН можно сформулировать набор требований к программному комплексу ИС «НИ РАН». Информационная система «Научный Институт РАН» должна:

- обеспечивать решение основных информационных задач научного института в составе Российской Академии Наук;
  - позволять гибко изменять конфигурацию системы под нужды конкретной организации, реализацию новых модулей для решения специфических задач;
  - предоставлять средства интеграции и структуризации существующих данных;
  - обеспечивать поддержку распределенного взаимодействия, в том числе со сторонними системами (через специализированные адаптеры, создаваемые отдельно).
-

Система должна включать:

- средства интеграции существующих данных;
- автоматизированные интерактивные средства структуризации и пакетной загрузки данных;
  - пользовательские и административные интерфейсы ввода новых данных и управления уже находящимися в системе данными;
  - систему (возможно распределенную) хранения данных;
  - систему безопасности, обеспечивающую аутентификацию пользователей и авторизацию доступа к ресурсам системы;
  - спецификации по разработке дополнительных модулей, обеспечивающих решение специфических задач научного института.

ИС «НИ РАН» [3, 4] представляет типовой собой программный комплекс автоматизации информационной деятельности научного института в составе Российской академии наук, обеспечения научной деятельности его сотрудников, взаимодействующий с другими информационными системами в составе ЕНИП. Разработанная платформа ИС «НИ РАН» предоставляет широкие возможности по конфигурированию под нужды конкретного научного института. Ядро всей системы составляют инфраструктурные службы. Они обеспечивают хранение, индексирование и поиск ресурсов, обеспечивают безопасность и взаимодействие между другими модулями. Базовые компоненты ИС «НИ РАН» обеспечивают выполнение самых общих информационных задач научного института – управление содержанием портала, организационной структурой, ведение сведений о публикациях и проектах сотрудников. Все действия конечный пользователь производит через веб-интерфейс. ИС «НИ РАН» представляет собой модульную расширяемую систему, решающую типовые информационные задачи научного института в составе РАН. Но реальные потребности таких организаций и их сотрудников зачастую бывают очень специфичными и относятся к узкой предметной области. Для удовлетворения таких нужд разрабатываются прикладные подсистемы, расширяющие функциональность типового решения в конкретных экземплярах. Архитектура НИ РАН представлена на рис. 3.

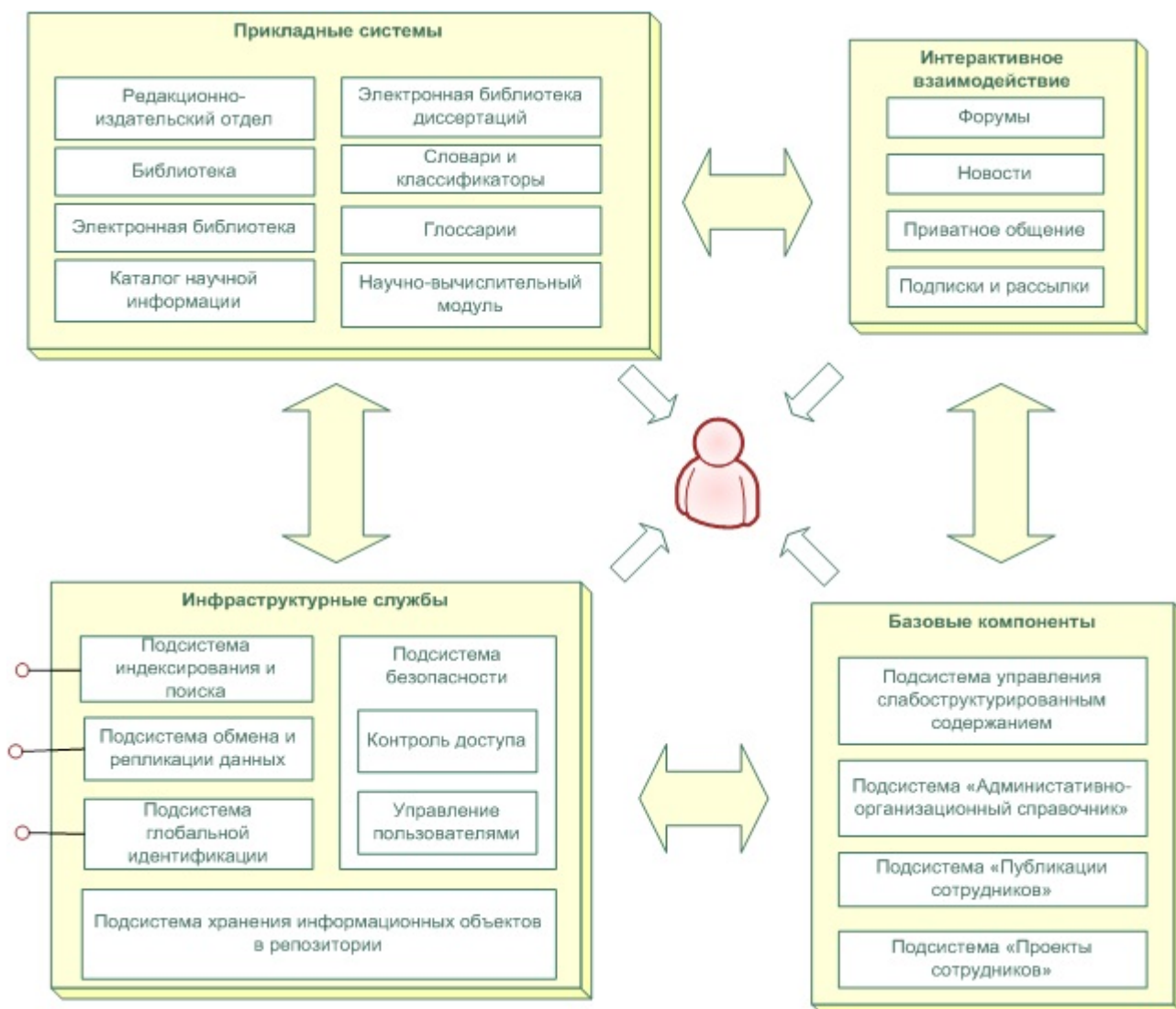


Рисунок 3. Архитектура НИ РАН

В основном профиле метаданных можно выделить общую поддержку следующих четырех основных групп информационных сущностей:

«Участники научной деятельности» – центральное звено, вся информация в РАН связана с научной деятельностью ее сотрудников, «Персон», образующих разнообразные организационные объединения от формальных («Организации» и «Подразделения») до неформальных («Коллективы», «Сообщества», «Рабочие группы»);

«Научная деятельность», в частности, «Проекты», отражающие процесс научной деятельности, информация о результатах проектов, патентах и т. п., а также «Научные мероприятия» – как разовые, так и повторяющиеся, такие, как «Конференции», «Семинары», «Симпозиумы»;



«*Результаты научной деятельности*», в которые могут входить «Интернет-системы» – вебсайты и пр., «Базы данных», предоставляющие автономные коллекции информации с той или иной степенью интеграции с ЕНИП и т. п., «Экспериментальные данные» и их «Математические модели», «Программные системы», в частности, «Научные вычислительные приложения», «Экспериментальные установки», «Изобретения», «Технологии» и т. п.

«*Документы и публикации*» – ресурсы этого типа представляют собой научные труды, статьи, отчеты сотрудников (научные «Публикации» и «Диссертации» сотрудников), возможно, административные «Постановления» и «Распоряжения». Примерами специализации публикации могут служить, например, «Тезисы конференций» и т. п. Профиль метаданных НИРАН представлен на рис. 4.

На базе информационной системы Научный институт РАН были созданы информационные системы ряда организаций (институтов и отделений) РАН:

- Отделение общественных наук (на базе системы Соционет);
- Библиотека по естественным наукам (БЕН РАН);
- Вычислительный центр (ВЦ РАН);
- Институт физики твердого тела;
- Палеонтологический институт;
- Пермский научный центр и Институт механики сплошных сред УРО;
- Институт проблем химической физики и Научный центр Черноголовка;
- Тихоокеанский океанологический институт им. В.И. Ильичева;
- Отделение математических наук (ОМН);
- Санкт-Петербургский научный центр;
- Дальневосточное отделение (ДВО РАН);
- Институт научной информации по общественным наукам (ИНИОН);
- Институт США и Канады (ИСКРАН);
- Институт проблем информатики (ИПИ РАН);

- Портал пространственных метаданных ГеоМета;
- Цифровая библиотека Научное наследие РАН;
- Северокавказский научный центр.

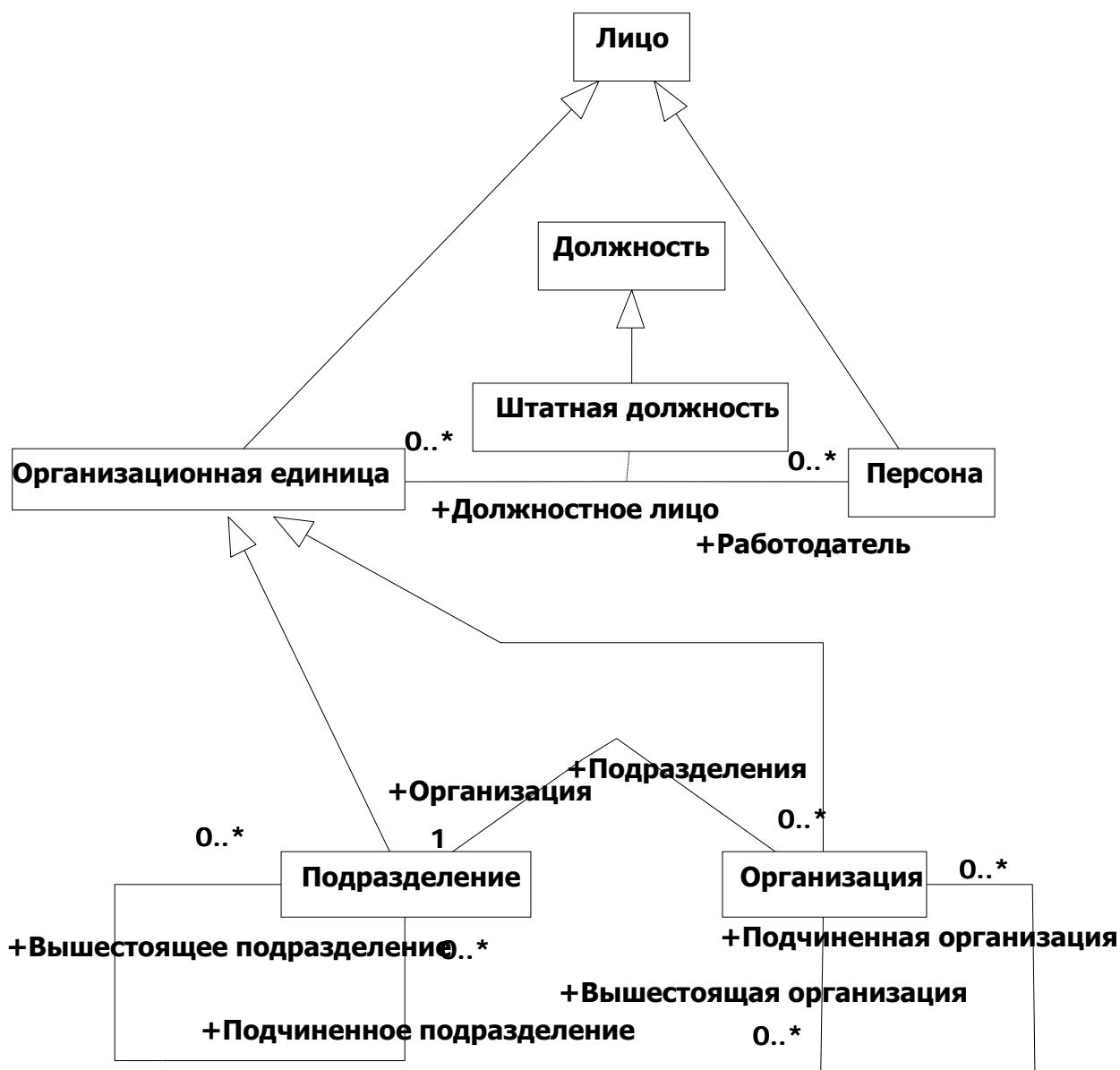




Рисунок 4. Основной профиль метаданных НИ РАН

### **3. ЕДИНОЕ НАУЧНОЕ ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО РАН**

Российская академия наук обладает уникальными научными информационными ресурсами. Среди них – опубликованные результаты научных исследований и экспериментов, библиографические и фактографические базы данных, сведения об ученых, их научной деятельности, публикациях, проектах и т. п. Эти ресурсы представляют значительный интерес для сотрудников РАН, членов мирового научного сообщества, для представителей промышленности и предпринимателей, которые заинтересованы во внедрении результатов научных исследований. Предполагалось, что Единое Научное Информационное Пространство (ЕНИП) РАН должно стать интегрированным источником научной информации.

Система предусматривала объединение сведений о разнородных научных информационных ресурсах РАН, обеспечение актуальности этих сведений и широких возможностей для достаточно точного поиска научных ресурсов на основе этих сведений, поддержку средств научной коммуникации, сервисов, связанных с возможностью оперативного информирования пользователей о необходимых им ресурсах, и т. п.

Такая система могла обеспечить пользователей актуальными данными о текущем состоянии и характеристиках информационно-научной базы институтов РАН и их подразделений, упростить анализ состояния и тенденций развития науки. Облегченный доступ к информации мог бы изменить способы ведения научной деятельности, способы обучения. Для обеспечения взаимодействия существующих разнородных научных систем в рамках ЕНИП предполагалось выработать корпоративные стандарты на интерфейсы взаимодействия, а также профили метаданных, что позволило бы реализовать инструментальные средства, обеспечивающие интеграцию данных в единую среду. Результатом решения этих первоочередных проблем должны были явиться предложения ЕНИП по:

- типовым интерфейсам взаимодействия (форматы данных, протоколы обмена) отдельных информационных источников (организаций РАН, поддерживающих собственные научные информационные ресурсы);
- профилям метаинформации, предоставляемой этими источниками; в частности, производится разработка набора элементов метаданных для научной информации общего характера, предложений по формированию элементов

метаданных для отдельных областей науки и согласование их с научным сообществом и международными открытыми стандартами;

- справочникам и классификаторам ресурсов;
- реализации политики информационной безопасности и требований по разграничению прав доступа к цифровым ресурсам.

Инициатива по организации Единого Научного Информационного Пространства (ЕНИП) РАН призвана была помочь научным коллективам сделать ряд шагов в направлении интеграции разнородных научных информационных и программных ресурсов отдельных научных учреждений, предоставлении пользователям более эффективных средств интеграции и поиска информации, научной коммуникации, сотрудничества и совместной работы. Под единым пространством понимается не формирование централизованной системы, не навязывание всем одних и тех же решений, а стремление последовательностью практических шагов, совместными усилиями научных коллективов:

- сформулировать взаимосогласованный набор соглашений, правил и открытых стандартов;
- приготовить совокупность макетов и типовых решений для реализации адаптеров прикладных систем, инфраструктурных служб, поддерживающих разные уровни интероперабельности распределенных гетерогенных данных и приложений;
- создать ряд информационных систем общего назначения, следующих этим соглашениям, использующих эти реализации, допускающих модульную организацию, наращивание функциональных возможностей;
- применить эти результаты для решения соответствующих задач научных учреждений.

Основу ЕНИП РАН составляют, прежде всего, стандарты на метаданные информации, циркулирующей в ЕНИП. Эти стандарты должны отвечать следующим требованиям:

- включать в себя основные типы информации, требующейся для поддержки работы научного сотрудника;
  - быть открытыми, т. е. обеспечивать доступ к соответствующей информации по этим описаниям;
-

- быть расширяемыми, т. е. обеспечивать возможность детализации описаний;
- обеспечивать возможности интеграции информации;
- обеспечивать возможности уникальной идентификации информации;
- обеспечивать возможности размещения и поиска информации в распределенной среде;
- быть ориентированными на современные и перспективные технологии описания и использования информации (в нашем понимании – ориентироваться на семантический Веб (Semantic Web));
- обеспечивать возможности интероперабельности с внешней средой.

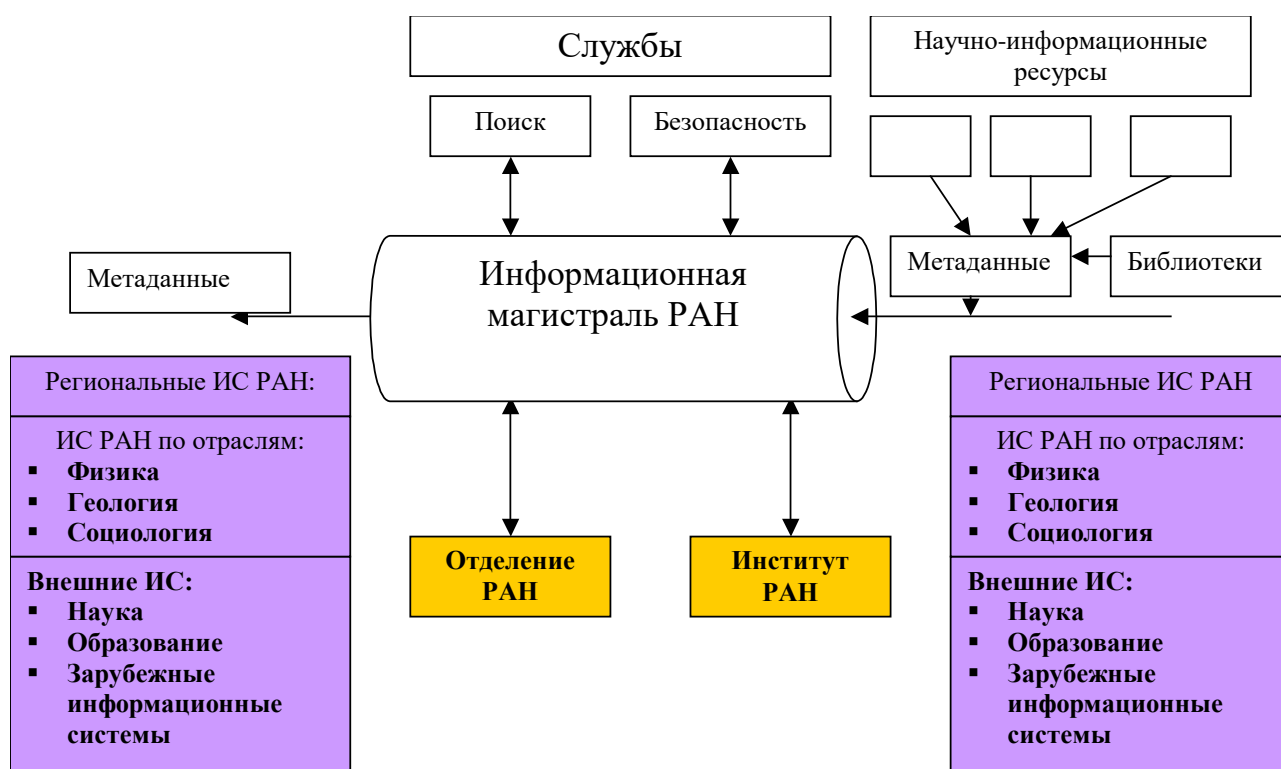


Рисунок 5. Информационная магистраль ЕНИП РАН

Основу единого информационного пространства РАН составляет Информационная магистраль ЕНИП РАН (рис. 5), представляющая собой комплекс аппаратных, программных и организационных мер, обеспечивающих:

- формирование состава цифровых ресурсов и служб ЕНИП РАН;
- предоставление доступа к цифровым ресурсам и службам ЕНИП РАН;
- обеспечение защиты цифровых ресурсов и служб ЕНИП РАН;

- ведение и поддержка в актуальном состоянии метаданных системы;
- поиск по хранимой метаинформации и идентификация ресурсов;
- интеграцию ресурсов различных областей и отраслей знаний.

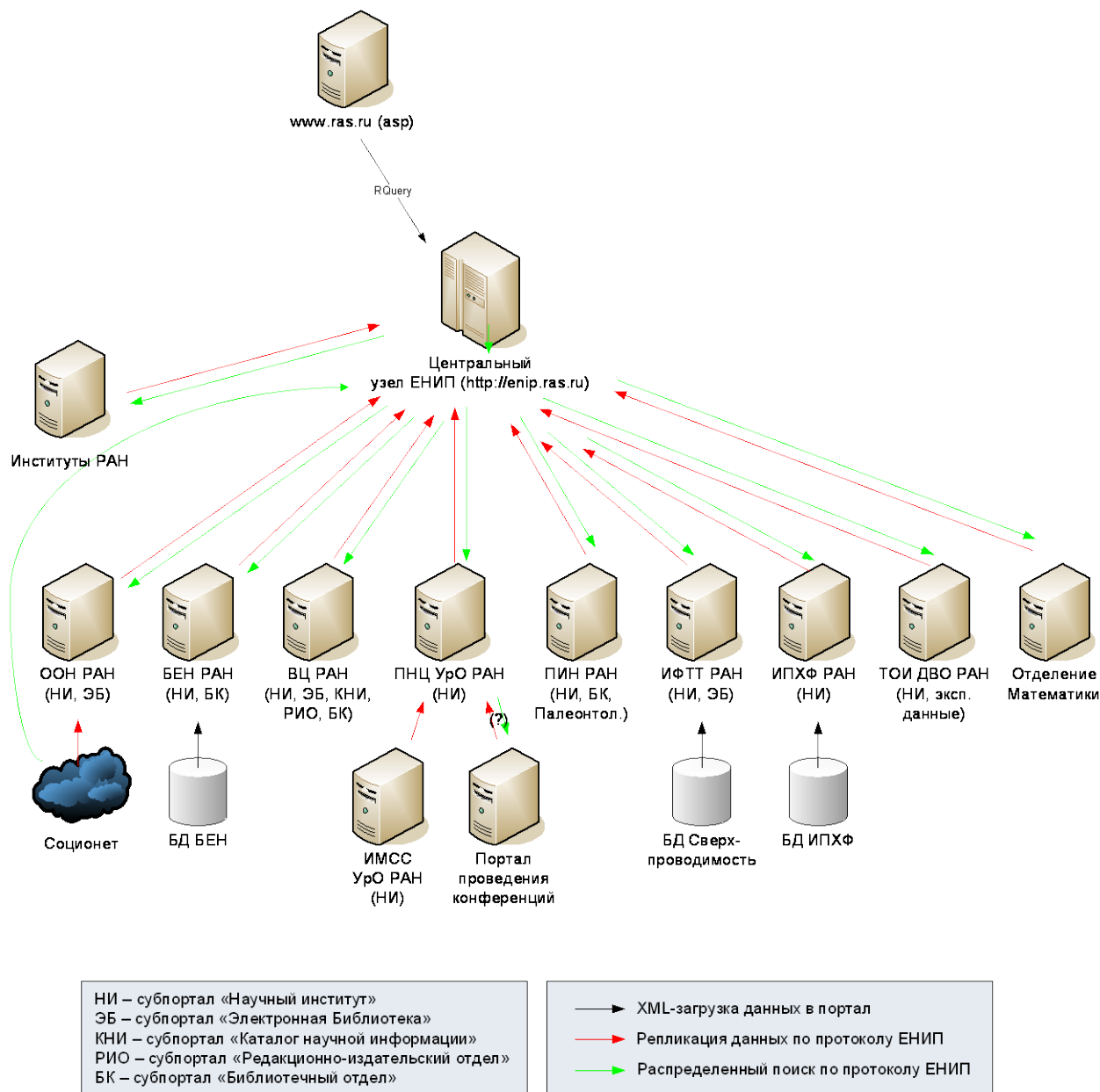


Рисунок 6. Схема взаимодействия узлов ЕНИП

На рис. 6 изображена схема взаимодействия узлов ЕНИП –четыре типа узлов ЕНИП: центральный узел; узлы организаций; независимые источники данных; независимые системы, включенные в ЕНИП.

Центральный узел осуществляет интеграцию данных с других узлов с помощью механизма репликации (копирования) метаинформации. На основе

реплицируемой на центральный узел метаинформации строятся поисковые индексы и на их основе осуществляется единый поиск по этим узлам. Загрузка данных в узлы системы может осуществляться из других источников, например, из сайтов организаций. Независимые информационные системы могут быть включены в ЕНИП самостоятельно, если обеспечены протоколы взаимодействия. Система ЕНИП оперирует такими ресурсами, как персоны, публикации, организации, подразделения и проекты. Данные по этим ресурсам обновляются каждую неделю. Центральный узел предоставляет пользователям две возможности поиска: поиск по локальной базе данных и полнотекстовый поиск. Поиск по локальной базе осуществляется по стандартным ресурсам: персона, организация, публикация, проекты. Актуальность информации может составлять разницу в 6 дней от информации на сервере-источнике данных. Полнотекстовый поиск позволяет получить полную и актуальную информацию, но выполняется дольше.

#### **4. ЦБ «НАУЧНОЕ НАСЛЕДИЕ РОССИИ»**

Учитывая важность формирования цифровых библиотек, Российская академия наук приняла в 2006 году целевую научную программу «Создание ЦБ «Научное наследие России»». ЦБ призвана аккумулировать цифровые копии книг, статей, документов, хранящихся в библиотеках, архивах и музеях РАН. В первую очередь акцент сделан на перевод в цифровую форму редких и уникальных изданий, важнейших документов по истории РАН, материалов экспозиционного характера, включая аудио-видеоматериалы [5]. Основной целью создания ЦБ является предоставление через интернет всем желающим информации о выдающихся российских ученых, внесших вклад в развитие фундаментальных естественных и гуманитарных наук, с возможностью ознакомления с полными текстами опубликованных ими наиболее значительных работ. Исходя из этой цели, в ЦБ было решено включать не только электронные версии книг, но и развернутые сведения о российских ученых – биографические данные, основные этапы их научной деятельности, разнородную архивную и музейную информацию, отсканированные фотографии, аудио- и видеозаписи, относящиеся к теме научного наследия.

Другой целью создания ЦБ является обеспечение сохранности оригиналов изданий, являющихся исторической ценностью, поскольку возможность работы с цифровыми копиями существенно снижает потребности в работе с печатными



материалами, а каждая «книговыдача» на руки раритетных изданий сокращает срок их «жизни».

Третьей целью создания ЦБ является включение сведений об ученых и основных результатах их научной деятельности в Единое научное информационное пространство (ЕНИП) РАН. На начальном этапе реализации основными задачами Программы были разработка основных принципов формирования ЦБ, технологии сканирования, обработки и предоставления пользователям материалов, включаемых в Библиотеку, а также создание программного обеспечения, сопровождающего все этапы создания ЦБ.

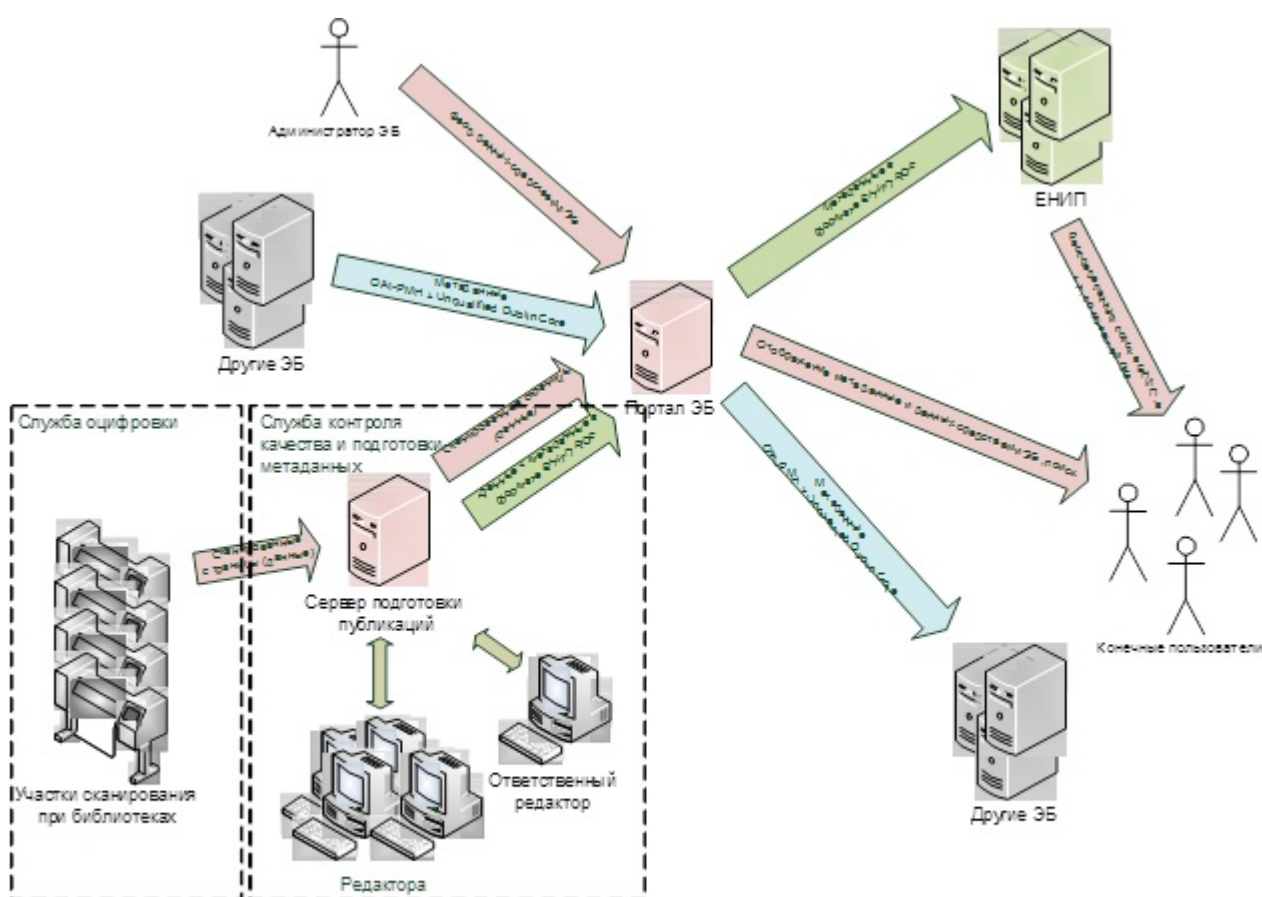
В основу технологии формирования ЦБ положен принцип распределенного наполнения и централизованной поддержки. Руководство Программой осуществляет Межведомственный суперкомпьютерный центр РАН, осуществляющий вместе с ВЦ РАН и БЕН РАН разработку технологии и программного обеспечения наполнения и поддержки ЦБ. Основными поставщиками информации для загрузки в ЦБ в настоящее время являются центральные академические библиотеки (БАН и БЕН РАН с их отделами в институтах и научных центрах РАН), ИНИОН, Центральный архив РАН с его Санкт-петербургским филиалом, Геологический музей РАН им. В.И. Вернадского, Институт русской литературы РАН (Пушкинский дом). В настоящее время наполнение ЦБ осуществляется копиями изданий, которые не подпадают под действие закона о защите авторских прав (в основном это издания, вышедшие из печати до 1920-го года). Основные элементы функциональности распределенной цифровой библиотеки следующие (рис. 7):

- доступ к ресурсам – запрос, определение местоположения, извлечение, трансформация и сохранение ресурса; поиск может осуществляться как по атрибутам ресурса, так и по полным текстам;
- управление ресурсом – создание нового ресурса, внесение его в ЦБ, удаление старого ресурса и изменение существующего;
- управление метаданными – их создание, обработка и преобразование; состав метаданных определяется соглашениями;
- управление словарями – их создание, обработка и преобразование; состав словарей определяется соглашениями;

- управление участниками – их регистрация, подписка, права доступа и персональная информация;
- управление цифровой библиотекой – управление коллекциями, группами пользователей, членством, так же, как общее управление политикой, качеством или функциональностью;
- системное администрирование – установка, конфигурирование, необходимые периодические мероприятия, восстановление после сбоев и мониторинг ЦБ.

Рисунок 7. Функциональная схема ЦБ «Научное Наследие»

Цифровая библиотека строится как распределенная информационная си-



стема с выделенным центральным узлом. Узлы системы, с одной стороны, являются точками входа в цифровые библиотеки организаций – участников проекта, с другой, – поставщиками информации для всей распределенной системы. Таким образом, ключевой принцип архитектуры – независимое развитие цифровых

библиотек организаций – участников с одновременной интеграцией данных в единое информационное пространство. Это достигается стандартизацией предоставления метаданных, форматов предоставления данных, интерфейсов поиска и словарей. Таким образом, каждая из цифровых библиотек организаций – участников может хранить данные в собственных форматах и предоставлять собственные сервисы, но в то же время должна обеспечить единые для всех интерфейсы, упомянутые выше.

Центральный узел системы должен обеспечить навигацию, поиск и предоставление данных по всем цифровым библиотекам в соответствии с унифицированными форматами и сервисами. Сервера хранения оцифрованных данных обеспечивают надёжное хранение и резервирование оцифрованных данных библиотеки, а также подмножества метаданных, отражающих структуру информации (например, оглавление книг). Кроме того, на них возлагается задача по предоставлению доступа к данным конечных пользователей, перенаправленных с центрального портала цифровой библиотеки. Серверами хранения данных для центров оцифровки предоставляются также средства автоматизации размещения и поддержания актуальности данных. Центральный веб-портал цифровой библиотеки «Научное Наследие РАН» осуществляет консолидацию метаданных, полученных из центров оцифровки, в рамках централизованного хранилища, обеспечивая, таким образом, централизованный доступ к ним пользователей. Взаимодействуя с серверами хранения оцифрованных данных, он является также единой точкой доступа к электронным версиям научных трудов. Второй задачей, решаемой центральным порталом, является обеспечение интеграции библиотеки в ЕНИП РАН путём предоставления на центральный сервер ЕНИП метаданных, по которым возможен распределённый поиск.

Функциональная схема ЦБ «Научное Наследие» приведена на рис. 7, главная страница – на рис. 8.

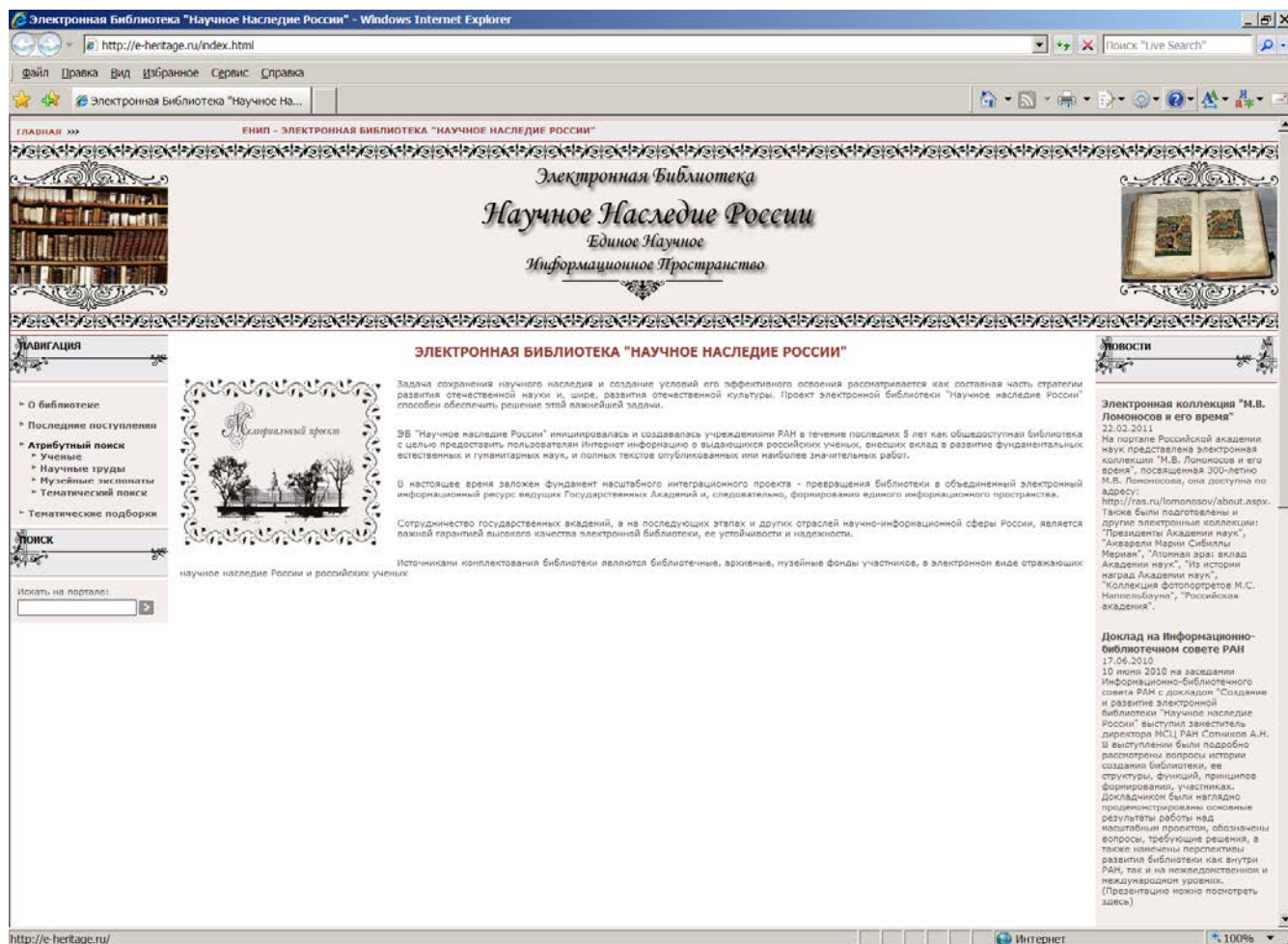


Рисунок 8. Главная страница ЦБ «Научное Наследие»

## 5. ПОРТАЛ ГЕОМЕТА

К настоящему времени в учреждениях РАН накоплен большой опыт использования геоинформационных технологий, реализованы многочисленные геоинформационные проекты, созданы базы и банки пространственных данных. Академические ресурсы пространственных данных составляют значительную часть национальных информационных ресурсов. Основным производителем пространственных данных являются учреждения геологического, геофизического, географического и экологического (природоохранного) профилей. В то же время данные рассредоточены, их использование ограничено зачастую рамками того проекта, где они созданы, затруднены или невозможны поиск существующих данных и доступ к ним, не налажен обмен ими. Причина этого – отсутствие эффективной системы управления пространственными данными. Ее создание позволит интегрировать данные и знания о территории, строить и использовать модели

природных и социально-экономических явлений и процессов, их взаимодействия в системе «общество – природная среда», использовать методы пространственного анализа, обеспечивать территориальное планирование и управление. В целом в учреждениях РАН имеется опыт выполнения разнообразных геоинформационных проектов для различных приложений, сформированы подразделения, отделы и лаборатории геоинформатики, укомплектованные высокопрофессиональными научными кадрами, располагающими необходимой технической базой, современными программными средствами геоинформационных систем (ГИС) и данными, то есть созданы необходимые условия для разработки ГИС и их интеграции.

Основным инструментом интеграции и предоставления пространственных данных в настоящее время являются геопорталы. Понятие «геопортал» означает точку входа в интернет с инструментами просмотра метаданных, поиска географической информации, ее визуализации, загрузки, распространения и, возможно, поиска геосервисов. Современное требование к системам поддержки геопорталов – независимость, расширяемость и гибкость компонентов, являющаяся важной особенностью современной программной системной архитектуры. Существует потребность в объединении этих данных, имеющих распределенный характер, в концептуально одну информационную систему, в обеспечении централизованного доступа к ним, в создании на основе интернета технологий единого информационного пространства геоданных.

Портал «ГеоМета» [6] – это стандартизированная и децентрализованная среда управления пространственной информацией, разработанная для доступа к базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников, облегчающая обмен пространственной информацией между организациями и ее совместное использование посредством интернета. Этот подход к управлению географической информацией имеет целью предоставить широкому сообществу пользователей средства для простого и своевременного доступа к имеющимся пространственным данным и существующим тематическим картам, которые могут оказаться полезными для поддержки информированного принятия решений. Главная цель портала – увеличить доступность разнообразных междисциплинарных данных различного масштаба вместе с

сопутствующей информацией, организованных и документированных стандартным и непротиворечивым способом, улучшить кооперацию и координацию усилий при сборе данных, сохраняющих ресурсы и, в то же самое время, ограждающих данные и информацию от нежелательного доступа.

Портал «ГеоМета» представляет собой платформу для создания распределенной среды интеграции неоднородных источников геоинформационных данных и предоставления к этой среде единой точки входа (веб-портала), которая позволит ученым в сфере наук о Земле легко находить специализированные данные и приложения, производить вычислительные эксперименты, визуализировать результаты деятельности.

Благодаря тому, что портал «ГеоМета» построен на базе ИС «НИ РАН» [1], являющейся базовым инфраструктурным компонентом ЕНИП [2], он может интегрироваться в ЕНИП с предоставлением расширения схемы геопространственными метаданными и геоданными. К функциональностям ГИС-части системы относятся:

- каталогизация, сбор, поиск геопространственных метаданных;
- размещение геоданных в собственном хранилище и предоставление к ним доступа;
- предоставление доступа к распределенным геопространственным данным по стандартизованным протоколам;
- визуализация карт, редактирование элементов.

Интерфейс системы представлен веб-порталом, поэтому для ГИС-части основным методом доступа пользователя к информации является обычный доступ к веб-страницам портала через любой распространенный браузер. Ядро системы предоставляет следующие возможности: управление статическим содержанием; хранение объектов системы (представленных RDF-тройками) в реляционных СУБД; индексирование и полнотекстовый поиск; обеспечение безопасности системы. Система поддерживает следующие основные типы ресурсов: *Пространственные данные* (картографические данные и их метаданные) и дополнительные типы ресурсов, такие, как *Организация, Персона, Публикация, Проект*, различные рубрикаторы и классификаторы.

Ресурс *Пространственные данные* содержит наборы пространственных данных и метаданные распределенных пространственных данных. Ресурс

---

Организация включает организации РАН, научные центры и другие организации. Данные об их сотрудниках сопоставлены ресурсу *Персона*. Ресурс *Проект* поддерживает сведения о проектах, выполненных или ведущихся в РАН и других ведомствах. Ресурс *Публикация* представляет данные о публикациях и научной деятельности.

Доступ к portalу осуществляется интерактивно через интернет посредством веб-браузера (например, Netscape Navigator или Microsoft Internet Explorer) по ссылке <http://www.geometa.ru>.

Главная страница портала представлена на рис. 9.

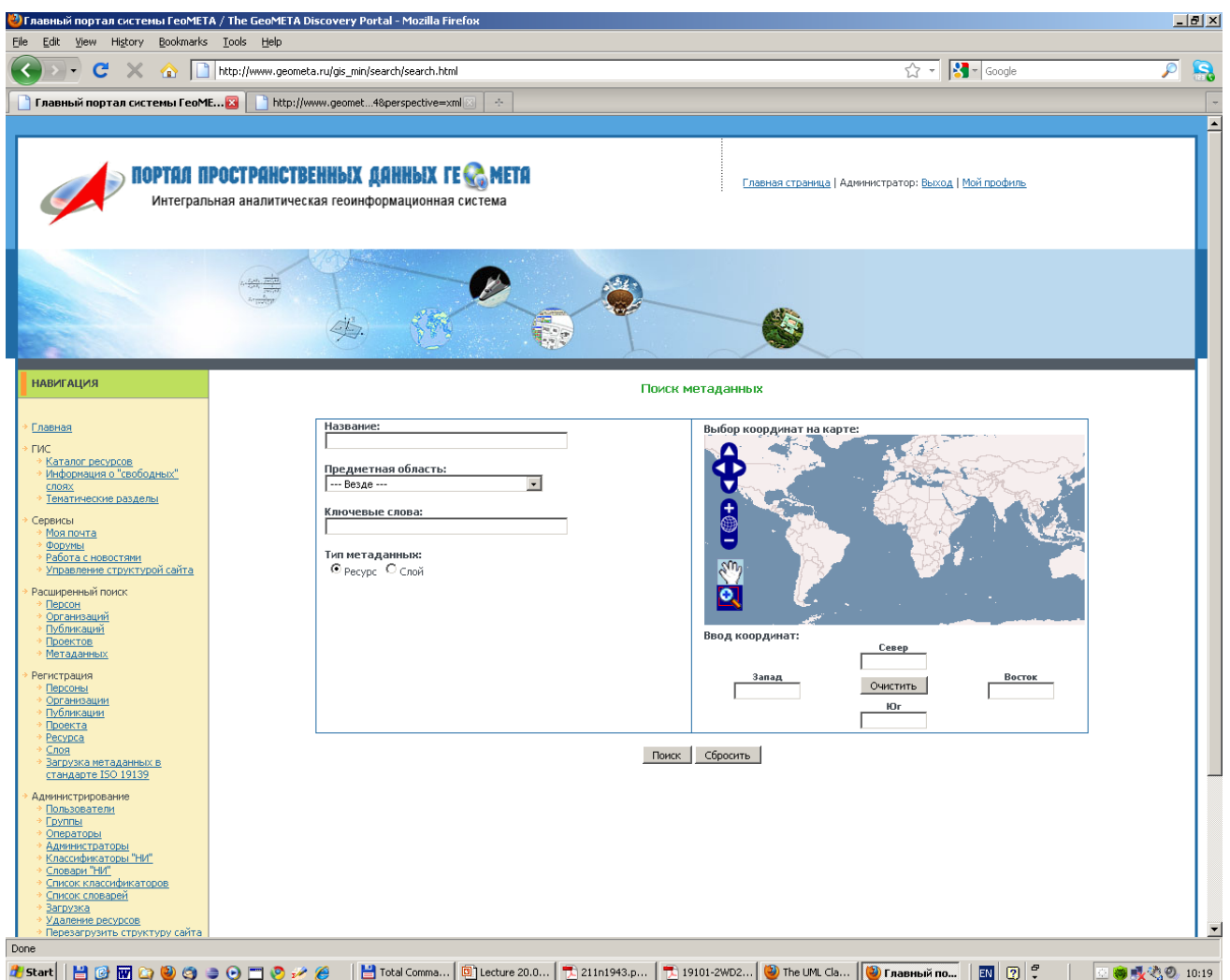


Рисунок 9. Главная страница портала ГеоМета

## **6. ПЕРСОНАЛЬНЫЕ СЕМАНТИЧЕСКИЕ ЦИФРОВЫЕ БИБЛИОТЕКИ**

Под персональными семантическими цифровыми библиотеками подразумеваются такие цифровые библиотеки, наполнение которых индивидуально для каждого пользователя системы и выполняется в полуавтоматическом режиме из разнородных источников данных, интегрированных в облако LOD. Будем далее для краткости называть их персональными открытыми цифровыми библиотеками или ПОЦБ. Типы информационных ресурсов и их структура определяются пользователем, исходя из своих интересов, то есть пользователь описывает интересующую его предметную область, определяя тематическое наполнение библиотеки.

Основная задача системы заключается в предоставлении пользователю унифицированного представления для возможности автоматизированного извлечения интересующей его информации по определенной предметной области. Представление ресурсов библиотеки в виде связанных данных расширяет функциональность семантических цифровых библиотек, давая возможность:

- включения дополнительных элементов описания данных информационных ресурсов;
- полного или частичного обновления данных из источников;
- использования интерфейсов для создания запросов к интегрированным в LOD источникам данных на основе SPARQL;
- включения в описания ресурсов других типов информации.

Одна из задач, которая решается в ПОЦБ, – это реализация интеграции набора данных в пространство LOD с использованием онтологии предметной области информационных ресурсов, т. е. автоматизированное обнаружение новых наборов данных и, по возможности, установка и поддержка связей с элементами данных из этих наборов данных с уже имеющимися ресурсами в репозитории библиотеки, обеспечивая одновременно рекомендуемую проектом LOD функциональность в рамках одной системы.

Источники данных подразделяются на два типа: внешние и внутренние. Внешними мы называем те источники, которые интегрированы в LOD, и данные которых представлены в RDF и доступны нам с использованием SPARQL. Для своих практических целей мы использовали такие известные источники в LOD, как DBpedia, Europeana. Внутренние источники могут представлять собой любой



другой тип источника данных, который не интегрирован в LOD. На практике в качестве внутренних источников мы использовали другие библиотеки, которые предоставляли доступ к своим данным по протоколу OAI-PMH.

К основной функциональности системы, реализующей ПОЦБ относятся:

- функции атрибутивного поиска;
- функция выделения неявных связей между ресурсами по их описаниям;
- функция работы с коллекциями;
- создание/просмотр/редактирование/объединение/вложенные коллекции;
- функция отображения онтологии ИД;
- функция детализации, которая обеспечивает преобразование в подзапросы, соответствующих различным ИД;
- функция для выполнения запросов и обработки результатов и предоставления окончательного результата пользователю;
- функция автоматического мониторинга ИД на наличие новых/измененных данных;
- создание словарей, классификаторов, тезаурусов;
- редактирование элементов;
- поддержка («гибкой») классификации ресурсов;
- поддержка настройки уровней доступа к различным ветвям тезауруса.

Исходя из определения источников данных ПОЦБ и перечня функций системы, можно выделить «внутренние» функции, т. е. те, которые оперируют данными в рамках системы и интегрируют данные из «внутренних» источников и фактически определяют обычную семантическую библиотеку. «Внешние» функции обеспечивают подключение и извлечение данных из LOD и позволяют задать тематическое наполнение библиотеки и установить связи, таким образом задавая фактически определение ПОЦБ.

Онтология ПОЦБ разработана в общем виде без привязки к конкретным методам и способам реализации семантических цифровых библиотек. Фактически общая онтология ПОЦБ состоит из двух онтологий:

1) онтология СЦБ, построенная на основе онтологии информационных систем, включающая в себя основные понятия, необходимые для обеспечения

основной функциональности библиотеки, такие, как ресурс, пользователь, коллекция, словарь, классификатор, запрос, источник и т. д.

2) онтология и тезаурус предметной области, для которой пользователь определяет ее понятия, их тип, структуру, совокупность словарей и классификаторов, которые представляют тезаурус предметной области, который обеспечивает доступ неквалифицированных пользователей, решающих задачи поиска информации, к знаниям предметной области в разных источниках. Эта онтология позволяет:

- выработать и зафиксировать общее понимание области знания;
- представить знания в удобном для обработки автоматизированными подсистемами виде, обеспечить возможность получения и накопления новых знаний, а также представить возможность многократного использования знаний.

Тезаурус же обеспечивает терминологическую поддержку и помогает пользователям сформулировать запрос к системе, в том числе, подобрать правильные ключевые слова для описания искомого результата, имеющихся данных и контекстной информации. Задача автоматизированного поиска релевантных источников данных осложняется тем, что чаще всего информация о связях между ними предоставляется в основном на уровне данных с помощью связей *sameAs*, *seeAlso*. Даже простой анализ связей *sameAs*, *seeAlso* на уровне найденных данных позволит выявить эквивалентные классы, ранее не определенные связи между разными источниками или новые источники. Описание связей на уровне схем затем можно использовать при формировании запросов к источникам данных.

До недавнего времени связи между источниками на уровне схем описывались гораздо реже. В последние несколько лет эта задача решается с введением и активным распространением спецификации VOID. Для описания источников RDF-данных, в которой предоставляется информация о связанных источниках данных. VOID-описание содержит информацию об используемых словарях, статистическую информацию о том, сколько ресурсов того или иного типа или значений определенных свойств используются во множестве. При создании словаря VOID была сведена к минимуму необходимость создания новых свойств и классов путем использования существующих словарей. Например, для описания статистической информации используется словарь SCOVO. На основе этой информации можно делать вывод о релевантности источника тому или иному запросу или

---

предметной области. В рассматриваемой системе VoID-описание набора данных в хранилище генерируется с помощью D2R Server. В сгенерированное описание не попадают информация о подключенных источниках данных и статистика по имеющимся с ними связям. Для включения этой информации были использованы правила, по которым осуществляется поиск связанных данных. Полученное описание в рамках используемой системы позволяет формировать распределенные запросы к подключенным источникам данных в терминах онтологии, используемой в этой системе. С использованием VoID-описаний запросы из системы транслируются в термины уже источников данных. Также это описание применяется для отображения обобщенного результата поиска.

На рис. 10 представлена общая схема подключения различных источников данных с использованием технологий из стека проекта LOD.

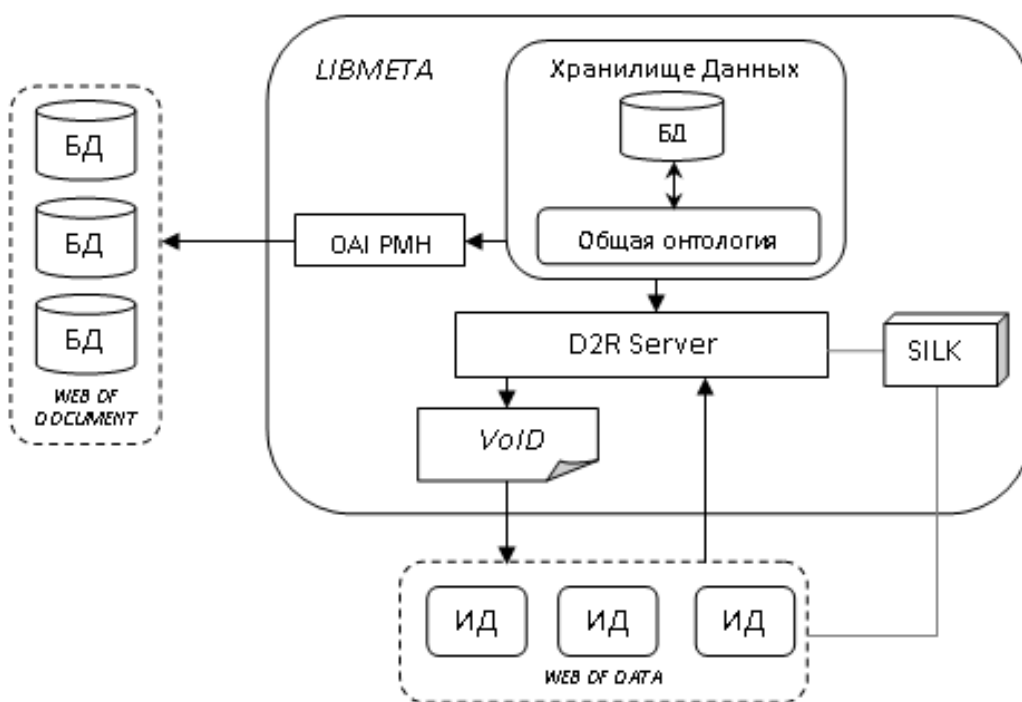


Рисунок 10. Схема подключения различных источников данных

Доступ к данным Libmeta осуществляется через ее общую онтологию, которая, как было сказано, состоит из: а) онтологии семантической библиотеки, б) онтологии предметной области, которая задает тематическое направление информационных ресурсов. При этом D2R Server использует онтологию Libmeta для

создания SPARQL-точки доступа к ее данным. Используются правила, которые задаются для каждого подключаемого источника (правил может быть несколько), с помощью которых осуществляются поиск и сохранение связей между данными Libmeta и источником из LOD. Для задания правил связывания используется фреймворк SILK. Правила описываются в соответствии с требованиями SILK и хранятся в определенном для каждого источника месте. После описания правила и указания его расположения все действия по запуску и анализу результатов работы SILK выполняются программно, для этого используется соответствующая задаче версия фреймворка.

При каждом подключении нового источника или обновлении набора связей уже подключенных нужно обновлять VOID-описание множества данных Libmeta, анализируя полученный набор ссылок и правила, по которым они выполнялись. Это позволит обновить статистическую и структурную части VOID, необходимых для использования при формировании запросов в терминах общей онтологии и их преобразования в запросы к релевантным источникам в соответствующим им терминах.

Libmeta также исторически поддерживает обмен данными по протоколу OAI-PMH с библиотеками, неинтегрированными в LOD, выступая агрегатором, который интегрирует их данные в LOD.

В рамках создания первой версии ПОЦБ был реализован проект по созданию стандартизированной и децентрализованной среды управления информацией электронных фондов Libmeta. В проекте реализованы средства интеграции приложений с разными источниками/каталогами метаданных/данных, сервис директорий метаданных, унифицированный интерфейс поиска данных.

Существенное различие во внутренних моделях данных, используемых в различных музеях, библиотеках и архивах, является главной проблемой на пути решения задачи интеграции данных. Для преодоления этой проблемы в решаемой задаче интеграции данных было предложено участникам экспортировать метаданные из своего внутреннего формата в формат на базе Dublin Core с использованием синтаксиса XML, так как во внутренних используемых форматах удастся выделить общую часть, которая ложится в рамки предложенного формата. В системе используется универсальный модуль загрузки метаданных в произвольном XML-формате в соответствии с протоколом OAI-PMH.

---

## ЗАКЛЮЧЕНИЕ

За прошедшие годы была проделана значительная работа по созданию информационных систем для обеспечения доступа к информационным ресурсам РАН и интеграции этих ресурсов. К сожалению, так и не была решена задача создания единого информационного пространства РАН, основанного на современном подходе к интеграции данных и приложений, и эта проблема остается в том же положении, что и была в конце 1990-х и начале 2000 гг.

## Благодарности

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проекты № 17-07-00214 и № 18-07-00297.

## СПИСОК ЛИТЕРАТУРЫ

1. Бездушный А.Н., Жижченко А.Б., Кулагин М.В., Серебряков В.А. Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек. // Программирование. 2000. №4. С. 12–20.

2. Интегрированная система информационных ресурсов: архитектура, реализация, приложения. Коллектив авторов под редакцией В.А. Серебрякова. М.: ВЦ АН, 2004. 240 с.

3. Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН. Вычислительный центр РАН. Москва, 2006, 238 с.

4. Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М., Теймуразов К.Б., Филиппов В.И. Информационная Web-система «Научный институт» на платформе ЕНИП. М.: Вычислительный центр им. А.А. Дородницына РАН, 2007. 257 с.

5. 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. №4. С. 28–31.

6. Атаева О.М., Кузнецов К.А., Серебряков В.А., Филиппов В.И. Портал интеграции пространственных данных «ГеоМета». Препринт ВЦ РАН, 2010. 106 с.

7. Атаева О.М., Серебряков В.А. Персональная цифровая библиотека Libmeta как среда интеграции связанных открытых данных// Электронные библиотеки: перспективные методы и технологии, электронные коллекции:

Всероссийская научная конференция RCDL-2014 (Дубна, 13–16 октября 2014 г.): труды конференции / сост. Л.А. Калмыкова, М.Р. Когаловский. Дубна: ОИЯИ, 2014, С. 66–71.

---

## **ELECTRONIC LIBRARIES IN THE COMPUTING CENTER OF RUSSIAN ACADEMY OF SCIENCES – MAIN DEVELOPMENTS**

V. A. Serebryakov

*Dorodnicyn Computing Centre of Federal Research Centre "Computer Science and Control" of Russian Academy of Sciences, Moscow 119333. Vavilov str., 40*

serebr@ultimeta.ru

### **Abstract**

The main projects that have been implemented in the Computing Center named A.A. Dorodnitsyna of the Russian Academy of Sciences (CC RAS) for the last 20 years, that is, since 1998, are analyzed. One of the first was the implementation of the pilot project "Integrated Information Resource System (ISIR) RAS". The successful completion of this project allowed the development of work on the integration of heterogeneous scientific information resources into the general academic scientific information system. An important stage was the project of creating the Unified Scientific Information Space (ENIP) of the RAS. This project was based on the subsystem "Scientific Institute of the Russian Academy of Sciences", created at the CC of the Russian Academy of Sciences and the Center for Scientific Telecommunications (CNTK) of the Russian Academy of Sciences. Considering the importance of building digital libraries, in 2006 the Russian Academy of Sciences adopted the target scientific program "Creating the Central Bank "Scientific Heritage of Russia", in accordance with which the digital library was implemented. The created GeoMeta Portal is a standardized and decentralized spatial information management environment designed to access geodatabases, map products and associated metadata from various sources, facilitating the exchange of spatial information between organizations and its sharing via the Internet. Currently, the main line of work is the LibMeta digital personal semantic library. The main task of this system is to provide the user with a unified view for the possibility of automated extraction of information of interest to him on a particular subject area.

---

**Keywords:** *subject area, scientific subject area, scientific information, scientific knowledge, generalized representation of scientific subject area, taxonomy, thesaurus, global ontology, search engines, organization of scientific knowledge, digital libraries*

## REFERENCES

1. *Bezdushnyj A.N., Zizhchenko A.B., Kulagin M.V., Serebryakov V.A.* Integrirovannaya sistema informacionnyh resursov RAN i tekhnologiya razrabotki cifrovyyh bibliotek // *Programmirovanie*. 2000. No 4. S. 12–20.
2. *Integrirovannaya sistema informacionnyh resursov: arhitektura, realizaciya, prilozheniya.* Kollektiv avtorov pod redakciej V.A. Serebryakova. M.: VCz AN, 2004. 240 s.
3. *Bezdushnyj A.A., Bezdushnyj A.N., Serebryakov V.A., Filippov V.I.* Integraciya metadannyh Edinogo Nauchnogo Informacionnogo Prostranstva RAN. Vychislitel'nyj centr RAN. Moskva, 2006, 238 s.
4. *Bezdushnyj A.A., Bezdushnyj A.N., Nesterenko A.K., Serebryakov V.A., Sysoev T.M., Tejmurazov K.B., Filippov V.I.* Informacionnaya Web-sistema «Nauchnyj institut» na platforme ENIP. M.: Vychislitel'nyj centr im. A.A. Dorodnicyna RAN, 2007. 257 s.
5. *Kalenov N.E., Savin G.I., Serebryakov V.A., Sotnikov A.N.* Principy postroeniya i formirovaniya ehlektronnoj biblioteki «Nauchnoe nasledie Rossii»// *Programmnye produkty i sistemy*. 2012. №4. S. 28–31.
6. *Ataeva O.M., Kuznecov K.A., Serebryakov V.A., Filippov V.I.* Portal integracii prostranstvennyh dannyh «GeoMeta». Preprint VC RAN, 2010. 106 s.
7. *Ataeva O.M., Serebryakov V.A.* Personal'naya cifrovaya biblioteka Libmeta kak sreda integracii svyazannyh otkrytyh dannyh// *Ehlektronnye biblioteki: perspektivnye metody i tekhnologii, ehlektronnye kollekcii: Vserossijskaya nauchnaya konferenciya RCDL-2014 (Dubna, 13–16 oktyabrya 2014 g.): trudy konferencii / sost. L.A. Kalmykova, M.R. Kogalovskij.* Dubna: OIYAI, 2014. S. 66–71.

**СВЕДЕНИЯ ОБ АВТОРЕ**



**СЕРЕБРЯКОВ Владимир Алексеевич**, гл. н. с. ВЦ РАН, д. ф.-м. н., профессор, окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

**Vladimir Alekseevich SEREBRYAKOV**, Chief Researcher of the Computing Center of the Russian Academy of Sciences, D. Sc., professor, graduated from M.V. Lomonosov Moscow State University. Specialist in the field of algorithmic languages and information technology.

e-mail: [serebr@ultimeta.ru](mailto:serebr@ultimeta.ru)

*Материал поступил в редакцию 10 декабря 2018 года*



УДК 004

## ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ЭКСПЕРТНОЙ ОЦЕНКИ ЗАЯВОК В КОНКУРСАХ ФЕДЕРАЛЬНОЙ ЦЕЛЕВОЙ ПРОГРАММЫ

С. М. Гарина<sup>1</sup>, Т. В. Модянова<sup>2</sup>

*Федеральное государственное бюджетное научное учреждение (ФГБНУ)  
«Дирекция научно-технических программ», 123557, Москва, ул. Пресненский  
Вал, 19, с. 1*

<sup>1</sup>garina@fcntp.ru, <sup>2</sup>modyanova@fcntp.ru

### **Аннотация**

Экспертиза в научно-технической сфере имеет более чем трехсотлетнюю историю и за это время во внешних своих формах претерпела существенные изменения. Компьютерные технологии повлияли на процессы организации экспертизы особенно сильно. В статье описана процедура организации экспертизы в конкурсах на получение финансирования в рамках федеральной целевой программы (ФЦП) «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы». Разработанные специализированные информационные системы и сервисы, использующие новые технологические средства поиска и обработки данных, дают возможность эксперту всесторонне и глубоко оценить объект экспертизы. Информационная поддержка охватывает все стороны организации экспертизы и деятельности экспертов. Процедура организации экспертизы, используемая в ФГБНУ «Дирекция научно-технических программ», и ее информационно-аналитическое сопровождение обеспечивают функционирование современной системы независимой научно-технической экспертизы, позволяют повысить качество оценки проектов и одновременно с этим снизить трудоемкость процесса организации экспертизы и, как следствие, организовывать экспертизу в больших объемах, в сжатые сроки и с минимальными издержками.

**Ключевые слова:** научно-техническая экспертиза, организация экспертизы, информационное обеспечение, конкурсный отбор, критерии оценки, экспертная оценка, эксперт, информационная система, информационный сервис.

## **ВВЕДЕНИЕ**

ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России» в разных версиях действует с 1996 года, она явилась в некотором роде продолжением Комплексных программ научно-технического прогресса, которые разрабатывались в СССР с середины 1970-х годов [1]. Государственным заказчиком ФЦП является Министерство науки и высшего образования Российской Федерации (далее – Минобрнауки). В 2013 году была утверждена ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы» (далее – Программа) [2].

Процесс реализации любой ФЦП включает наполнение программных мероприятий конкретными проектами. Отбор проектов в соответствии с российским законодательством должен осуществляться путем проведения конкурсов. Неотъемлемым условием конкурсного отбора является проведение экспертизы, что существенно повышает правильность и обоснованность выбора наиболее перспективных проектов для финансирования. И хотя до сих пор отсутствует правовая норма об обязательности экспертизы, на практике она имеет широкое распространение в реализации различных ФЦП [3].

«Неотъемлемой составляющей механизма реализации Программы является использование современной системы независимой экспертизы на всех стадиях реализации Программы, позволяющей отбирать наиболее перспективные проекты для государственного финансирования, проводить независимую объективную экспертизу конкурсных заявок в целях выявления победителей, предложивших лучшие условия выполнения проектов, и осуществлять эффективную экспертную проверку качества полученных результатов. Экспертиза и отбор проектов в структурообразующих блоках Программы имеют сквозной характер и основываются на принципах объективности, компетентности и независимости. В рамках реализации Программы также обеспечивается проведение мониторинга финансируемых проектов» [2].

Организация экспертизы в соответствии с паспортом Программы возложена на ФГБНУ «Дирекция научно-технических программ» (далее – Дирекция) – специализированный орган оперативного управления программами и проектами

---

Минобрнауки. В Дирекции накоплен огромный опыт организации экспертизы для нескольких ФЦП и других механизмов бюджетного финансирования: экспертизы предложений при формировании тематики будущих конкурсов; заявок на получение финансирования; качества полученных результатов по исполняемым контрактам. При современном уровне развития компьютерных технологий организация экспертизы немыслима без информационных средств поддержки как деятельности эксперта, так и собственно организации экспертизы, поэтому внедрению современных средств обработки информации, работы с документами и других аспектов автоматизации процессов экспертизы постоянно уделяется пристальное внимание.

В рамках настоящей статьи, не претендуя на полноту охвата накопленного опыта, мы опишем сложившуюся в Дирекции процедуру организации и информационного сопровождения научно-технической экспертизы заявок, участвующих в конкурсах на получение финансирования в рамках Программы.

#### **ПРОЦЕДУРА ОРГАНИЗАЦИИ ЭКСПЕРТИЗЫ КОНКУРСНЫХ ЗАЯВОК**

Можно выделить следующие составляющие, присущие любому процессу организации экспертизы в научно-технической сфере [4]:

- критерии, по которым проводится оценка объекта экспертизы;
- состав вводной информации для экспертизы;
- качественный и количественный состав экспертов;
- метод получения экспертного мнения;
- способ обработки экспертных мнений и формирование итогового ранжирования;
- регламент проведения экспертизы в целом и сбора экспертных мнений в частности;
- методическое, организационное, информационное и техническое сопровождение процессов проведения экспертизы.

Рассмотрим каждый из этих аспектов применительно к организации экспертизы заявок, участвующих в конкурсном отборе на получение финансирования в рамках Программы.

**Критерии, по которым проводится оценка объекта экспертизы.** Один из этапов подготовки предстоящего конкурсного отбора – это формирование набора

показателей для оценки участвующих в конкурсе проектов. Критерии конкурсного отбора проектов по мероприятиям Программы утверждены Научно-координационным советом Программы: стоимость проекта, качество проекта, квалификация исполнителей, опыт исполнителей. Данные критерии задают общую структуру оценки заявок в рамках Программы. Чтобы отразить специфику конкурса, внутри каждого критерия формируются его показатели. В экспертизе заявок применяется количественный метод оценки, поэтому для каждого показателя устанавливается максимальное значение в баллах, в ряде случаев показатели могут иметь неравнозначный вес при подсчете итоговой экспертной оценки проекта. Раздел конкурсной документации «Оценка заявок на участие в конкурсе», где изложены показатели, их содержание и максимальное количество баллов по каждому показателю, фактически является экспертной анкетой, с которой работает эксперт при оценке заявок.

**Состав вводной информации для экспертизы.** Каждый показатель в экспертной анкете предполагает наличие информации, на основании которой эксперт может дать по нему балльную оценку. Конкурсная документация готовится таким образом, чтобы участник конкурса, подготавливая пакет документов для участия в конкурсе, предоставил необходимую информацию для оценки по каждому показателю. Для удобства заявителей и с целью оптимизации структуры и содержания заявок был создан информационный *Портал регистрации заявок* <http://konkurs2014.fcpir.ru>, который, используя стандартные формы конкурсной документации, позволяет сформировать заявку в электронном виде и затем перевести ее на бумажный носитель.

**Качественный и количественный состав экспертов.** Создание группы экспертов – один из ключевых этапов в процессе организации экспертизы. Основой для составления экспертных групп, которые будут привлечены к экспертизе, является реестр экспертов Программы, утвержденный Минобрнауки. Соответствующая база данных, содержащая сведения об экспертах, постоянно пополняется и актуализируется. База данных хранится в *Системе экспертиз* <https://sstp.ru> и через интегрированный программный модуль позволяет подбирать экспертов, квалификация и специализация которых соответствует объекту экспертизы. Число экспертов, привлекаемых к оценке каждого альтернативного проекта, определяется балансом между требованиями к достоверности экспертных оценок и

---

ограничением средств на оплату труда экспертов; по оценкам авторов статьи [5], в группе должно быть 3–5 экспертов, что соответствует сложившейся практике организации экспертизы в рамках Программы.

Если в конкурсе одновременно участвует много заявок, то одному эксперту не под силу изучить и оценить все проекты. Количество проектов, которое может оценить один эксперт, зависит от многих параметров: индивидуальные качества эксперта, его квалификация, объем пакета документов, который необходимо изучить, время, отведенное на экспертизу, и др. В литературе встречаются исследования, согласно которым число одновременно оцениваемых альтернатив не должно превышать девяти [5]. В рамках Программы заявки по близким темам объединяются в группы не более 10–12 проектов. Каждая группа проектов через *Систему экспертиз* <https://sstp.ru> передается на экспертизу специалистам в соответствующей области знаний, при этом автоматически контролируется отсутствие аффилированности экспертов с участниками конкурса.

Проведение экспертизы с использованием информационной системы *Система экспертиз* <https://sstp.ru> позволяет систематизировать информацию по работе с экспертами, просматривать историю проведенных экспертиз, что помогает сотрудникам, организующим экспертизу, осуществлять качественный подбор экспертов.

**Метод получения экспертного мнения.** Экспертиза осуществляется в режиме удаленного доступа через *Систему экспертиз* <https://sstp.ru>. Эксперты получают доступ к документам, которые подготовили авторы заявки, а также ко вспомогательным информационным ресурсам, которые позволяют эксперту получить дополнительную информацию об объекте экспертизы. К таким ресурсам относятся: текстовый документ «*Справка по направлению исследований заявки*», информационная система «*Система интеграции источников научно-технической информации РУСНТИ*», информационный сервис «*Проверить копирование*». «*Справка по направлению исследований заявки*» готовится на основе информации из наукометрических баз данных и предоставляет эксперту набор количественных показателей, которые характеризуют, насколько разработка, представленная в заявке, соответствует мировым трендам развития науки и техники. В информационной системе РУСНТИ агрегирована информация из многих

источников научно-технической информации и реализован интеллектуальный поиск семантически близких объектов, что позволяет находить проекты, близкие по тематике к объекту экспертизы, с целью избежать повторного финансирования. Информационный сервис «Проверить копирование» совмещает в себе несколько известных алгоритмов поиска похожих документов и позволяет выявлять случаи прямого копирования текстов. Перечисленные специализированные информационные системы и сервисы дают возможность эксперту всесторонне и глубоко оценить объект экспертизы.

Изучив содержание проектов, представленных на конкурс, и дополнительную информацию о них, эксперты заполняют электронную форму экспертной анкеты в *Системе экспертиз* <https://sstp.ru>. Вопросы анкеты представляют собой набор показателей по каждому критерию оценки с возможностью выбора балльной оценки из предельно допустимого количества баллов по данному показателю в соответствии с конкурсной документацией. Эксперт дает обоснование или пояснение по каждой выставленной балльной оценке. Иногда по желанию заказчика эксперту предлагается выбор из списка готовых вариантов обоснований. После ответа на все вопросы анкеты система осуществляет автоматический подсчет итоговой балльной оценки, и эксперт делает общий вывод в пользу или против финансирования данного проекта. По завершении работы с экспертной анкетой создается файл в формате .pdf – экспертное заключение в виде, удобном для чтения и обработки.

**Способ обработки экспертных мнений и формирование итогового ранжирования.** После того, как получено необходимое количество экспертных оценок по всем заявкам, проводится ранжирование проектов на основании полученных оценок. Существуют разные способы обработки экспертных мнений [5–9]. В тех случаях, когда мнения экспертов имеют количественное выражение, усреднение является наименее затратным по времени и материальным ресурсам способом построения итогового рейтинга, хотя этот способ имеет известные недостатки: игнорируется неравнозначность баллов разных экспертов, не учитывается возможная вариация средних оценок, особенно если эксперт оценивает не все объекты.

Результаты экспертных оценок в виде итогового рейтинга передаются в конкурсную комиссию Минобрнауки России, которая принимает решение о победителях. В *Системе экспертиз* реализовано рабочее место члена конкурсной

комиссии, которое обеспечивает доступ к документам конкурсных проектов и к экспертным заключениям.

**Регламент проведения экспертизы в целом и сбора экспертных мнений в частности.** Регламент проведения экспертизы определяется нормативными документами Минобрнауки и внутренними документами Дирекции. В них определены цели, объекты и субъекты экспертизы, принципы организации и проведения экспертизы, порядок проведения экспертизы, организация работы с экспертами.

Регламент работы с экспертами определяет принципы формирования общего списка экспертов Программы и принципы отбора экспертов для привлечения к проведению экспертизы конкретного объекта. В соответствии с регламентом с каждым экспертом заключается договор, в котором обозначены стоимость и сроки экспертизы, оговорены права и обязанности сторон, порядок передачи материалов и приемки оказанных услуг.

**Методическое, организационное, информационное и техническое сопровождение процессов проведения экспертизы.** Методическое сопровождение проведения независимой экспертизы заключается в выработке общих подходов к подбору экспертов и проведению экспертизы, подготовке и актуализации нормативно-методической базы, включающей:

- разработку методических рекомендаций проведения оценки объектов экспертизы по критериями оценки, утвержденным Заказчиком;
- разработку и актуализацию форм экспертных анкет и экспертных заключений;
- совершенствование состава документов, представленных в объектах экспертизы;
- систематизацию данных об объектах экспертизы;
- анализ данных о результатах выполненной оценки;
- рекомендации Заказчику по корректировке критериев оценки проектов;
- консультационную поддержку экспертов в процессе проведения экспертизы.

Организационное сопровождение и консультационную поддержку осуществляют кураторы экспертов – сотрудники Дирекции, которым поручены взаимодействие с экспертами, подбор экспертов, анализ результатов полученных

экспертных оценок, участие в разработке нормативно-методической базы и ряд других функций.

Информационная поддержка в виде информационных систем и сервисов, разработанных по заказу Дирекции, сопровождает все этапы конкурса – от формирования заявки для участия в конкурсе до составления итогового рейтинга [10].

Техническое сопровождение, в случае необходимости, осуществляют сотрудники организации-разработчика программного обеспечения информационных систем и сервисов.

### **ЗАКЛЮЧЕНИЕ**

Описана процедура экспертизы заявок на участие в конкурсах на получение финансирования в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», перечислены информационные системы и сервисы, разработанные с целью совершенствования процесса экспертизы. Актуальная информационная поддержка процессов экспертизы позволяет обеспечить качественное и эффективное функционирование современной системы независимой научно-технической экспертизы и дает возможность организовать экспертизу в сжатые сроки, в больших объемах и с минимальными издержками.

Сложившаяся система независимой научно-технической экспертизы, разработанная в ФГБНУ «Дирекция научно-технических программ», прошла успешную апробацию при участии в реализации ряда федеральных и ведомственных целевых программ, а также внепрограммных мероприятий в научно-технической сфере. Методическое и информационно-аналитическое сопровождение постоянно совершенствуются, включают новые знания, технологии и опыт в области экспертного оценивания.

### **СПИСОК ЛИТЕРАТУРЫ**

1. Варшавский А.Е. Комплексные программы научно-технического прогресса страны – успешный пример реализации индикативного планирования. Анализ и моделирование экономических и социальных процессов // Математика. Компьютер. Образование. 2014. Т. 21. № 2. С. 185–197.



2. Постановление Правительства РФ от 21.05.2013 г. № 426 «О федеральной целевой программе «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы».

3. Плетнев К.И., Лазаренко Н.Е. Экспертиза в научно-технической сфере: методология и организация. М., РАГС, 2003.

4. Рыбаков Ю.Л., Голубев В.П., Дивуева Н.А., Медведев В.И., Ефимов Б.И. Обзор существующих в научно-технической сфере экспертных технологий (из опыта работы отечественных экспертных систем) // *Инноватика и экспертиза: научные труды*. 2012. № 2. С. 173–182.

5. Бухарин С.Н., Гукасов В.М., Лазаренко Н.Е. Теоретические и методические основы экспертизы фундаментальных и прикладных научно-технических проектов // *Инноватика и экспертиза: научные труды*. 2011. № 2. С. 58–66.

6. Орлов А.И. Экспертные оценки. Учебное пособие. М., 2002.

7. Гехман А.В., Якунин Ю.Ю., Даничев А.А., Володин А.А. Обработка результатов экспертиз в реестре научно-технических разработок // *Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева*. 2010. № 6 (32). С. 30–34.

8. Бухарин С.Н., Дивуева Н.А., Марышев Е.А. Выбор результирующего ранжирования в процессе научно-технической экспертизы инновационных проектов // *Инноватика и экспертиза: научные труды*. 2014. № 1. С. 114–120.

9. Губанов Д.А., Коргин Н.А., Новиков Д.А., Райков А.Н. Сетевая экспертиза. М.: Эгвес, 2011. 163 с.

10. Гарина С.М., Власов Д.В., Петров А.Н., Шуртаков К.В. Опыт построения информационной среды для решения задач реализации ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014-2020 годы» // *Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск)*. — М.: ИПМ им. М.В. Келдыша, 2018. С. 125–135. URL: <http://keldysh.ru/abrau/2018/proc.pdf>

---

## INFORMATION SUPPORT OF EXPERT EVALUATION OF APPLICATIONS IN COMPETITIONS OF THE FEDERAL TARGET PROGRAM

S. M. Garina<sup>1</sup>, T. V. Modyanova<sup>2</sup>

*Directorate of State Scientific and Technical Programs, 123557, Moscow, Presnensky Val, 19, building 1*

<sup>1</sup>garina@fcntp.ru, <sup>2</sup>modyanova@fcntp.ru

### **Abstract**

Peer review in the scientific and technical field has more than three hundred years of history and during this time its external forms has changed significantly. Computer technology has influenced the processes of the peer review organizing particularly strongly. The article describes the procedure for the peer review in tenders for funding under the federal target program "Research and development in priority areas of development of the scientific and technological complex of Russia for 2014–2020." The worked out specializing information systems and services using new technological tools for data retrieval and processing enable the reviewer evaluate the peer review's object fully and thoroughly. Information support covers all aspects of the peer review organizing and the expert's activities. The procedure for peer review organizing, established at the Directorate of Scientific Technical Programs, and its informational and analytical support ensure the functioning of an independent scientific and technical examination system and allow to organize peer review in large quantities, in a short time and with minimal costs.

**Keywords:** peer review organizing, information support, competitive selection, evaluation criteria, expert assessment, expert, reviewer, information system, information service.

### **REFERENCES**

1. Varshavskij A.E. Kompleksnye programmy nauchno-tekhnicheskogo progressa strany – uspešnyj primer realizacii indikativnogo planirovaniya. Analiz i modelirovanie ehkonomicheskikh i social'nyh processov // Matematika. Komp'yuter. Obrazovanie. 2014. T. 21. № 2. S. 185–197.
2. Postanovlenie Pravitel'stva RF ot 21.05.2013 g. № 426 «O federal'noj celevoj programme «Issledovaniya i razrabotki po prioritetyim napravleniyam razvitiya nauchno-tekhnologicheskogo kompleksa Rossii na 2014–2020 gody».

3. Pletnev K.I., Lazarenko N.E. Ekspertiza v nauchno-tehnicheskoy sfere: metodologiya i organizaciya. M.: RAGS, 2003.

4. Rybakov YU.L., Golubev V.P., Divueva N.A., Medvedev V.I., Efimov B.I. Obzor sushchestvuyushchih v nauchno-tehnicheskoy sfere ehkspertnyh tekhnologij (iz opyta raboty otechestvennyh ehkspertnyh sistem) // Innovatika i ehkspertiza: nauchnye trudy. 2012. № 2. S. 173–182.

5. Buharin S.N., Gukasov V.M., Lazarenko N.E. Teoreticheskie i metodicheskie osnovy ehkspertizy fundamental'nyh i prikladnyh nauchno-tehnicheskikh proektov // Innovatika i ehkspertiza: nauchnye trudy. 2011. № 2. S. 58–66.

6. Orlov A.I. Ekspertnye ocenki. Uchebnoe posobie. M., 2002.

7. Gekhman A.V., Yakunin Yu.Yu., Danichev A.A., Volodin A.A. Obrabotka rezul'tatov ehkspertiz v reestre nauchno-tehnicheskikh razrabotok // Vestnik Si-birskogo gosudarstvennogo aehrokosmicheskogo universiteta im. akademika M.F. Reshetneva. 2010. № 6 (32). S. 30–34.

8. Buharin S.N., Divueva N.A., Maryshev E.A. Vybor rezul'tiruyushchego ran-zhirovaniya v processe nauchno-tehnicheskoy ehkspertizy innovacionnyh proektov // Innovatika i ehkspertiza: nauchnye trudy. 2014. № 1. S. 114–120.

9. Gubanov D.A., Korgin N.A., Novikov D.A., Rajkov A.N. Setevaya ehkspertiza. M.: EHgves, 2011. 163 s.

10. Garina S.M., Vlasov D.V., Petrov A.N., Shurtakov K.V. Opyt postroeniya informacionnoj sredy dlya resheniya zadach realizacii FCP «Issledovaniya i razrabotki po prioretnym napravleniyam razvitiya nauchno-tehnologicheskogo kompleksa Rossii na 2014-2020 gody» // Nauchnyj servis v seti Internet: trudy XX Vserossijskoj nauchnoj konferencii (17–22 sentyabrya 2018 g., g. Novorossijsk). M.: IPM im. M.V. Keldysha, 2018. S. 125–135. URL: <http://keldysh.ru/abrau/2018/proc.pdf>

## **СВЕДЕНИЯ ОБ АВТОРАХ**



**ГАРИНА Светлана Михайловна** – кандидат физико-математических наук, доцент, главный специалист Отдела организации экспертизы ФГБНУ «Дирекция научно-технических программ»;

**Svetlana Mikhailovna GARINA** – Candidate of Physical and Mathematical Sciences, Associate Professor, Chief Specialist of the Expertise Organization Department of the Directorate of Scientific and Technical Programs;  
email: garina@fcntp.ru



**МОДЯНОВА Татьяна Викторовна** – кандидат политических наук, заместитель руководителя Отдела организации экспертизы ФГБНУ «Дирекция научно-технических программ»,

**Tatyana Viktorovna MODYANOVA** – Candidate of Political Sciences, Deputy Head of the Department for Organization of Expertise of the Directorate of Scientific and Technical Programs;  
email: modyanova@fcntp.ru

*Материал поступил в редакцию 12 декабря 2018 года*