

**Труды XVII Всероссийской научной конференции**

# **Научный сервис в сети Интернет**

**Новороссийск, 21-26 сентября 2015 г.**

**ИПМ им.М.В.Келдыша  
2015**










УДК 519.7  
ББК 73  
Н 34

Н 34 Научный сервис в сети Интернет: труды XVII Всероссийской научной конференции (21-26 сентября 2015 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2015. — 336 с.  
ISBN 978-5-98354-015-6


Сборник содержит доклады, сделанные участниками XVII Всероссийской научной конференции «Научный сервис в сети Интернет». Конференция проходила с 21 по 26 сентября 2015 года. Подробную информацию о конференции можно найти по адресам <http://agora.guru.ru/abrau2015/> и <http://keldysh.ru/abrau/2015/>.












Конференция поддержана Российским фондом фундаментальных исследований, проект 15-07-20607-г.












## ОГЛАВЛЕНИЕ<sup>1</sup>










<b>А.Л. Александров</b> 	Три кита тестирования – процессы, метрики, риски .....	7
<b>О.М. Атаева, М.В. Кулагин, В.А. Серебряков</b> 	Основные понятия для построения формальной модели семантических библиотек и описания процессов интеграции в ней.....	8
<b>Д.Ю. Ахметов, А.М. Елизаров, Е.К. Липачёв</b> 	Информационные системы и сервисы комплексной поддержки периодических научных изданий .....	16
<b>В.А. Бахтин, О.Ф. Жукова, Н.А. Катаев, А.С. Колганов, В.А. Крюков, Н.В. Поддерюгина, М.Н. Притула, О.А. Савицкая, А.А. Смирнов</b> 	Использование интернета для обучения параллельному программированию..	26
<b>А.С. Борисенкова, Ю.С. Корухова</b> 	Об одном подходе к определению авторства музыкальных произведений.....	34
<b>Д.А. Варламов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова</b> 	Информационно-вычислительная система по кристаллохимии и минералогии <a href="http://www-mincryst">www-mincryst</a> : 18 лет научному Интернет-инструментарии .....	42
<b>В.А. Васенин, С.А. Афонин, А.С. Козицын, Д.Д. Голомазов</b> 	Система «ИСТИНА» для подготовки принятия решений на основе анализа наукометрической информации.....	51
<b>В.М. Волохов, Д.А. Варламов, А.В. Волохов, А.И. Прохоров, Г.А. Покатович</b> 	Построение среды компьютерного моделирования нанокompозитных материалов на базе комплекса вычислительных сервисов и высокоуровневых веб-интерфейсов .....	63
<b>Е.А. Гаврилина, М.А. Захаров, А.П. Карпенко, Е.В. Смирнова</b> 	Количественная оценка метакомпетенций учащегося на основе анализа его поведения в социальных сетях .....	69

---

<sup>1</sup>  — гиперссылка на презентацию к докладу.

<b>М.П. Галанин</b> 	Математическое моделирование многомерных квазистационарных электромагнитных полей в канале электродинамического ускорителя .....	78
<b>М.П. Галанин, А.П. Лотоцкий, А.С. Родин</b> 	Исследование применимости моделей упругопластического тела для описания высокоскоростного деформирования ленты из алюминия.....	79
<b>М.П. Галанин, В.В. Лукин, А.С. Родин</b> 	Интегрированная программная платформа сопровождения вычислительного эксперимента Теметос: архитектура и задачи.....	85
<b>М.М. Горбунов-Посадов, А.В. Ермаков, Т.А. Полилова</b> 	Мультимедийные иллюстрации в научной публикации.....	96
<b>А. М. Елизаров, Н.Г. Жильцов, А.В. Кириллович, Е.К. Липачёв</b> 	Семантическое аннотирование в системе управления физико-математическим контентом.....	98
<b>Н.М. Ершов, Н.Н. Попова</b> 	Разработка онлайн-репозитория эволюционных алгоритмов оптимизации....	104
<b>А.А. Захаров, А.В. Бойко, С.В. Степанов</b> 	Безопасное SaaS облако для проектирования технологических процессов добычи нефти .....	108
<b>А.Д. Изаак, О.Г. Мисюрина, Ю.А. Пупырев, Д.Е. Чебуков</b>	Индексация видеозаписей научных мероприятий в общероссийском математическом портале Math-Net.Ru .....	114
<b>А.В. Карпов, А.С. Деникин, А.П. Алексеев, В.В. Самарин, М.А. Науменко, В.А. Рачков</b> 	Анализ характеристик распада атомных ядер в рамках сетевой базы знаний NRV по низкоэнергетической ядерной физике.....	119
<b>Ф.О. Каспаринский</b> 	Использование программ ассоциативного картирования для управления распределенными информационными ресурсами .....	127
<b>Ф.О. Каспаринский</b> 	Кросспостинг в авторском информационном континууме.....	135
<b>Ф.О. Каспаринский, Е.И. Полянская</b> 	Интернет-активность учёного в рамках парадигмы инфоцентризма .....	141

<b>Ф.О. Каспаринский, Е.И. Полянская</b> 	
Интернет-представительства научно-образовательных организаций и проектов на специализированных сайтах и в социальных сетях: SWOT-анализ .....	150
<b>М.Р. Когаловский, С.И. Паринов</b> 	
Научная информационная система со встроенной таксономией научных отношений .....	159
<b>А.С. Козицын, С.А. Афонин, Д.Д. Голомазов, М.А. Занчурин, В.С. Астапов</b> 	
Автоматизация составления отчетов о научной деятельности .....	169
<b>К.А. Косолапов, В.А. Серебряков, К.Б. Теймуразов, О.Н. Шорин</b>	
Алгоритм слияния дублетных библиографических записей .....	173
<b>К.В. Кузнецов, В.М. Михелев</b> 	
Построение параллельного алгоритма индексирования с использованием fractal trees .....	181
<b>Г.С. Курганская</b> 	
Поведение агентов в облаке интернет-образования .....	187
<b>Е.М. Лаврищева, Л.Е. Карпов, А.Н. Томилин</b> 	
Системная поддержка решения бизнес-задач в глобальной информационной сети .....	193
<b>Ю.П. Липунцов, В.А. Серебряков</b> 	
Соотношение информационных и экономико-математических моделей .....	219
<b>И.А. Мбого, Д.Е. Прокудин, А.В. Чугунов</b> 	
Развитие процессов автоматизации при формировании информационного пространства поддержки междисциплинарных научных исследований .....	232
<b>Г.М. Михайлов, Ю.П. Рогов, А.М. Чернецов</b>	
Организация внешнего почтового smtp-сервера в научной организации .....	237
<b>Г.М. Михайлов, К.П. Беляев, А.Н. Сальников, Н.П. Тучкова, И. Кирхнер</b> 	
Учет данных наблюдений для анализа динамических процессов в океанской модели циркуляции .....	240
<b>Е.И. Моисеев, А.А. Муромский, Н.П. Тучкова, К.Б. Теймуразов</b> 	
Модели управления колебательными процессами и их терминологическое описание в информационной среде .....	244
<b>А.В. Никешин, Н.В. Пакулин, В.З. Шнитман</b> 	
Мутационное тестирование сетевых протоколов с использованием формальных моделей .....	259

<b>С.И. Паринов</b> 	
Открытая наука.....	267
<b>Д.В. Петров, В.М. Михелев</b> 	
Геоинформационные технологии анализа границ карьеров рудных месторождений .....	279
<b>А.А. Печников</b> 	
Сравнение трех проектов вебметрического ранжирования научных организаций.....	285
<b>А.И. Привезенцев, А.Ю. Ахлестин, Н.А. Лаврентьев, Н.Н. Лаврентьева, А.З. Фазлиев</b> 	
Онтологическое представление количественной спектроскопии.....	291
<b>А.И. Прохоров, Д.А. Варламов, Е.С. Амосова, М.Е. Соловьева, В.Е. Туманов, П.К. Берзигияров</b> 	
Инфраструктура веб-сервиса интеллектуальной проблемно-ориентированной системы научной аналитики по физической химии радикальных жидкофазных реакций в Интернет.....	298
<b>В.А. Серебряков, А.О. Еркимбаев, В.Ю. Зицерман, Г.А. Кобзев, К.Б. Теймуразов, Р.И. Хайрулин</b> 	
Систематизация физико-химических данных. Возможности онтологий и баз данных, принцип дополнительности.....	303
<b>А.В. Чугунов</b> 	
Исследование "электронного участия" с использованием методов и технологий автоматизированного мониторинга веб-ресурсов.....	311
<b>Д.М. Шакирова, Д.Ш. Сулейманов, О.А. Невзорова, М.И. Курманбакиев</b> 	
Личность и эпоха: виртуальное представление научного наследия.....	321
<b>М.В. Энгель, В.В. Белов</b> 	
Интегрированная информационная система тематической обработки данных дистанционного зондирования Земли .....	329

# ТРИ КИТА ТЕСТИРОВАНИЯ – ПРОЦЕССЫ, МЕТРИКИ, РИСКИ

А.Л. Александров

*Luxoft*

В работе рассматриваются проблемные области, связанные с подготовкой и проведением тестирования программного обеспечения при разработке заказных систем. Автором выбраны наиболее критичные с его точки зрения области.

Во-первых, это оценка трудозатрат на тестирование в проектах сопровождения, характеризующаяся невозможностью применения стандартных методик, используемых для оценки трудозатрат типа, например, СОСОМО или метод функциональных точек.

Во-вторых, это выбор и применение типовых метрик тестирования. Важность выбора правильных метрик определяется тем, что результаты измерений служат достоверной основой для принятия управленческих решений при выполнении проекта.

Наконец, в третьих, это риски, связанных с процессами тестирования. Это традиционно непопулярная тема (почему-то считается, что в тестировании рисков не бывает), пренебрежение которой привело к неудаче многих проектов. В работе приведены примеры таких рисков и рекомендации по их преодолению.

# ОСНОВНЫЕ ПОНЯТИЯ ДЛЯ ПОСТРОЕНИЯ ФОРМАЛЬНОЙ МОДЕЛИ СЕМАНТИЧЕСКИХ БИБЛИОТЕК И ОПИСАНИЯ ПРОЦЕССОВ ИНТЕГРАЦИИ В НЕЙ

О.М. Атаева<sup>1</sup>, М.В.Кулагин<sup>2</sup>, В.А. Серебряков<sup>1</sup>

*1 ВЦ РАН, Москва, 2 МСЦ РАН, Москва*

## **Введение**

Цифровые библиотеки воспринимаются обычными пользователями как электронные версии каталогов традиционных библиотек, которые содержат описания физических объектов библиотеки (обычно книг или других печатных изданий). Определение тематики, содержания и структуры объектов рассматривается и воспринимается ими как дополнительная, но не обязательная функция таких библиотек.

Развитие современных технологий также подталкивает к переопределению понятия контента библиотеки, в качестве которого не обязательно могут выступать традиционные описания печатных изданий, но и любые другие типы объектов [2], например, мультимедийные объекты: видео, звук, фотографии, музейные экспонаты, коллекции минералов, архивные материалы и многое другое.

Интернет также может рассматриваться как библиотека, стихийно наполняемая его пользователями без видимого порядка и структуры. Каждая страница имеет собственный идентификатор в виде URL и относится к некоторому объекту реального мира и таких страниц может быть тысячи. Чаще всего автоматическая обработка таких страниц для выявления полезной информации и различных взаимосвязей является сложной задачей и предполагает большие трудозатраты на извлечение нужной информации по узкой предметной области.

## **Предпосылки**

Для облегчения процесса автоматической обработки информации в сети некоторое время назад была предложена концепция «Linked Open Data» [5] для размещения и описания данных, опирающаяся на уже имеющиеся наработки парадигмы Semantic Web [4]. Основные принципы «Linked Open Data» призывают использовать для идентификации реальных ресурсов и их цифровых описаний, а также их взаимосвязей, уникальные идентификаторы URI. При этом эти URI должны обеспечивать доступ к описаниям объектов по протоколу HTTP и представлять описание в виде RDF [6]. Это обеспечивает возможность автоматической обработки информации, содержащей в своем описании ссылки в виде URI на другие взаимосвязанные ресурсы и их описания. Следуя этим принципам, обеспечивается стандартизированный механизм доступа к данным, и поддерживается глобальный обмен данными независимо от модели данных



конкретного узла, обеспечивая возможность их интеграции основываясь на URI. Парадигма Semantic Web позволяет структурировать описания ресурсов и представлять их в виде RDF, основываясь на онтологиях [7].

Итак, мы рассматриваем библиотеки как хранилища структурированных разнообразных данных с возможностью их интеграции в LOD и возможностью определения их тематической направленности. При этом подразумевается, что поддерживается вся традиционная для электронных библиотек функциональность: создание, редактирование, поиск, идентификация ресурсов. В дальнейших работах мы планируем представить онтологию такой библиотеки, архитектуру приложения и описать реализацию прототипа этого приложения. При этом мы пока опускаем описание процессов взаимодействия пользователя с этой библиотекой, а сосредоточимся на ее основных понятиях, которые необходимы для описания модели данных и процессов поступления (обогащения) и поиска данных для цифровой библиотеки определенной тематики, интегрированной с различными источниками данных.

### **Основные понятия**

Семантические библиотеки активно развиваются на протяжении последнего десятилетия [14], но их понятийная структура не является устоявшейся и активно развивается. Концептуальная модель электронных библиотек с определениями важнейших представлений об архитектуре, ресурсах и функциональности электронных библиотек была определена в программном документе «DELOS Digital Library Reference Model» [10]. С другой стороны, эффективность исследований в этой области зависит от стандартизации и формализации собственно описаний ресурсов таких библиотек и процессов их представления.

Далее будут даны определения для построения формальной модели любых типов ресурсов рассматриваемой семантической библиотеки, которые лежат в основе ее построения.

### **Роль тезауруса в семантической библиотеке**

Определяя информационные ресурсы, наборы атрибутов и информационные объекты, мы позволяем формировать понятийную основу конкретной предметной области для семантической библиотеки. Тезаурус же в свою очередь обеспечивает терминологическую поддержку понятий, помогая решать разнообразные задачи.

Одна из таких актуальных задач – возможность поддержки классификации ресурсов в библиотеке. Зачастую направлений классификаций бывает несколько. В некоторых случаях классификаторы могут динамически меняться в процессе жизнедеятельности системы. Более того, со временем могут потребоваться новые классификаторы, которые должны в идеале безболезненно и с минимальными трудозатратами внедряться в библиотеку, не требуя при этом внесения изменений в систему и не затрагивая структуры

данных хранилища системы. При этом часто возникают ситуации, когда классификацию необходимо бывает применить к различным типам ресурсов, причем классифицируемые типы ресурсов могут эволюционировать также по ходу развития системы. Мы называем такую поддержку процедуры классификации «гибкой классификацией»

### Основные определения

#### 1) Информационный объект

$O_B = \{o_1, \dots, o_n\}$  – множество информационных объектов,

#### 2) Информационный ресурс

$R = \{r_1, \dots, r_m\}$  – множество информационных ресурсов (типов информационных объектов),

a)  $TYPE(o) = r$  – каждому объекту соответствует информационный ресурс

b)  $IsRe(r_1, r_2)$  – определяет иерархию ресурсов, т.е.  $r_2$  подресурс  $r_1$ , ресурс может иметь 0 и более подресурсов

#### 3) Атрибуты

$A = \{a_1, \dots, a_k\}$  – множество атрибутов информационных ресурсов

##### a) Тип атрибутов

$A_D$  – множество атрибутов, значением которых может являться любое значение из некоторого множества  $D$ , в котором не может быть информационных ресурсов системы

$$\forall a_i, a_i \in A_D : Z(a_i) = d, d \in D, D \subset O$$

$A_I$  – множество атрибутов, значением которых может являться любой объект из  $I_O$

$$\forall a_i, a_i \in A_I : Z(a_i) = o, o \in O_B$$

$$A_D \vee A_I = A, A_D \wedge A_I = \emptyset$$

##### b) Вид атрибутов $V = \{v_1, \dots, v_l\}$

Атрибут  $a$  может быть нескольких видов

$$\forall a, a \in A : VIEW(a) = V', V' \subset V$$

##### c) Для атрибутов определены область определения $DOM(a) = R'$ , $R' \subset R$ и область определения $RAN(a) = X : X \subset D$ или $X \subset O_B$

#### 4) Наборы атрибутов

$SA$  – множество таких подмножеств атрибутов, в каждое из которых входят хотя бы по одному атрибуту каждого вида, каждое такое подмножество называется набором атрибутов

$\forall v, A' \ v \in V, A' \subset SA \ \exists a, a \in A' \ \text{VIEW}(a)=V', \ v \in V'$

$SET(r) = A''$  соответствие информационного ресурса некоторым наборам атрибутов  $A'' = \{A'\}, A' \subset SA$ ,

Если информационному ресурсу  $r$  соответствует некоторый набор атрибутов  $A' = \{a_1, \dots, a_i\}$ , то все объекты  $o \in O_B$ , такие что  $TYPE(o) = r$ , как представляются как набор  $(a_1, \dots, a_i)$

В соответствии с типом значений атрибуты могут быть однозначными, многозначными, имеющими множество однотипных значений, которые могут составлять мультимножество, множество, список, массив. Для описания формальной общей картины эти тонкости пока можно опустить, полагая, что мы можем включить в набор атрибутов столько однотипных атрибутов, сколько требуется.

5) *Модель данных источника:*

$$S_X = \{(r_X, A_r)\},$$

где  $r_X$  и  $A_r$  ресурсы и их атрибуты представленные в этом источнике,

6)  $G$  – функция отображения элементов модели данных источника на информационные ресурсы системы и их набор атрибутов,

$$G(r_X, A_r) = (r, A'), \ SET(r) = A', \ A' \subset SA$$

7) *Источник данных:*

$$X = (S_X, G)$$

8) *Тезаурус предметной области:*

$$Th = (P, T, R)$$

$P$  – набор понятий,

$T$  – набор вербальных терминов для понятий;

$R$  – набор горизонтальных и вертикальных связей между понятиями и терминами.

9)  $F$  – функция сопоставления каждому информационному объекту из  $O_B$  соответствующих терминов из  $T$

$$F(o_i) = \{t\}, \ t \in T, \ o_i \in O_B$$

10) *Качество данных* – это набор характеристик  $Q = (q_1, \dots, q_m)$

$DQ$ – функция, оценивающая информационные объекты на соответствие условиям характеристик качества и ставящая ему в соответствие некоторое число  $k$  из диапазона от 0 до 1, где значения близкие к 1 считаются более качественными;

$DQ(o_i) = k = \sum q_j(o_i)/m$ , если объект  $o_i$  удовлетворяет характеристике  $q_j$  то  $q_j(o_i) = 1$ , в противном случае  $q_j(o_i) = 0$

Основываясь на этих определениях, формулируется набор аксиом для рассматриваемой задачи, на основе которых строится, например, представление объединенных данных об объектах информационного ресурса для пользователя из разных источников:

$$\exists o_i : TYPE(o_i) = r \wedge r \in R \wedge SET(r) = A' \wedge G_1^{-1}(r, A') = S_{X1} \wedge \dots \wedge G_n^{-1}(r, A') = S_{Xn}$$

Запросы, предназначенные для нескольких источников, возвращают множество пар  $\{(o_i, S_x)\}$  информационных объектов и соответствующих им источников данных. В некоторых случаях пользователь может захотеть увидеть консолидированный объект, сформированный из результатов поиска, и в таком случае решаются в самом простом виде задачи очистки данных, такие как удаление дублирующих значений, объединение пересекающихся значений и т.д. Изменение структуры данных делать не приходится, так как эта задача решается на уровне возврата результатов с помощью функций  $G_i^{-1}$ . Даже из такого поверхностного обзора понятно, что если ранее задача приведения и очистки явно не обозначалась, этот процесс является неизбежным этапом при работе с любыми источниками данных.

### Качество данных

Качество данных является важной характеристикой любого набора данных независимо от его структуры. Опираясь на оценку качества данных, можно давать объективную оценку эффективности процессов, происходящих в семантической библиотеке, наиважнейшим из которых является интеграция данных с другими источниками, не обязательно являющимися библиотеками. Исходя из определения атрибутов и их видов при решении, например, задач связывания данных из разных источников, необходимо предварительно оценить как минимум качество значений идентифицирующих атрибутов. В противном случае, объективно проанализировать полученные результаты не представляется возможным. Поэтому используем формальное определение качества, удовлетворяющее нас в контексте нашей задачи.

Если данные пригодны для тех целей, для которых они предназначены, то можно сказать, что эти данные качественные [13]. Качество данных определяется связанными характеристиками, каждая из которых определяет наличие тех или иных ошибок во множестве данных [12].

При определении понятия *ошибка* будем использовать обозначения:  $o_1, o_2 \in O_B$ , где  $TYPE(o_1) = r$ ,  $TYPE(o_2) = r$ ,  $a_i$  атрибут из набора  $SET(r) = A'$ .  $val(Z(a_{1i}), Z(a_{2i})) = true$  – если значения атрибутов  $a_{1i}$  и  $a_{2i}$ , объектов  $o_1$  и  $o_2$  совпадают,  $f(a_j) = true$ ,  $i \leq n$ , если задан формат значений  $a_j$ , которые могут быть представлены в виде последовательности символов из непустого конечного

алфавита  $L$ , т.е., для таких значений  $a_j$  существует грамматика  $G_J$ , которая задает значения атрибута  $a_j$ ,  $U=\{u_1, \dots, u_n\}$  – множество условий, накладываемых на объекты,  $u(a_i, \dots, a_j)$ , будем обозначать как  $g(o_i, u_j)$  – функцию, принимающую значения  $true$  в случае выполнения условия  $u_j$  для объекта  $i$ , в противном случае,  $false$ .

Будем говорить о наличии на уровне данных [12, 13]

a) *синтаксических ошибок* (обозначим через  $M_{syn}$ ), если для

- $Z(a_i) = null$ , при этом не существует условия вида  $u(a_i, a_j)$ , когда от  $a_i$  нет зависимых значений
- $Z(a_i) \notin RAN(a_i)$
- $f(a_j) \neq true, a_j \in A_D$  и  $a_j \notin L(G_J)$

b) *семантических ошибок* (обозначим через  $M_{sem}$ ), если выполняется, что

- $Z(a_i) = null$ , при этом существует условие вида  $u(a_i, a_j)$ , такое что  $Z(a_j) \neq null$
- $g(a_i, u_j) \neq true, a_j \in A$
- если для любого  $o_1$ , существует  $o_2 \in O_B$  и  $val(o_1, o_2) = \wedge val(Z(a_{1i}), Z(a_{2i})) = true$

c) и о наличии ошибок на уровне схем данных [2, 3, 4, 7], если

$G(r_x, A_r) \neq (r, A')$ , то есть не существует функции отображения ресурсов источника на ресурсы системы

Определяем основные характеристики качества на уровне данных, такие как

- *точность*;
  - *синтаксическая точность*;
  - *семантическая точность*;
- *полнота*;
  - *синтаксическая полнота*;
  - *семантическая полнота*;
- *уникальность*.

Таким образом определены ошибки только двух типов, и каждый тип ошибок разбивается на подтипы в зависимости от того, какому критерию качества они соответствуют.

### **Зависимость уровня качества данных от порядка обработки ошибок**

Уровень качества данных определяется степенью соответствия рассматриваемых данных характеристикам, используемым при определении понятия «качества данных». Наилучшее качество достигается при полном соответствии всем характеристикам.

Для достижения высокого уровня качества данных, при приведении данных прежде следует обрабатывать синтаксические ошибки, а затем семантические. По определению синтаксических ошибок при их обработке происходит анализ отдельных значений атрибутов объектов независимо от других значений объекта и их связей. Из определения семантических ошибок при обработке следует анализ значений атрибутов, зависящих от значений других атрибутов объекта или от накладываемых ограничений, или на основании значений атрибутов происходит анализ зависимости различных объектов друг от друга. Поэтому синтаксическая корректность значений влияет на качество обработки семантических ошибок, так как часть ошибок может оказаться «невидимой».

### **Дальнейшая работа**

Сформулированы формальные определения для описания контента семантической библиотеки, являющегося ядром любой библиотеки. Мы исходили из того, что структура данных библиотеки будет определяться не на этапе проектирования библиотеки, а в процессе ее эксплуатации. При этом важной составляющей нашей библиотеки является возможность интегрирования данных с разными источниками. Соответственно, процесс интеграции данных является одним из ключевых процессов. На этапе интеграции основную роль играет функция  $G$ , которая отвечает за преобразование данных источника.

Понятийная модель ресурсов библиотеки, описанная в формальных терминах, позволяет на их основе построить онтологию с поддержкой тезауруса предметной области, что является на наш взгляд, необходимым условием для создания гибких семантических библиотек, интегрируемых в LOD и использующих все возможные преимущества взаимосвязанных данных из его источников.

### **ЛИТЕРАТУРА**

1. В. А. Серебряков, О. М. Атаева. Персональная цифровая библиотека Libmeta как среда интеграции связанных открытых данных // RCDL – 2014
2. А.Б. Антопольский, А.А. Каленкова, Н. Каленов, В.А. Серебряков, А. Сотников. Принципы разработки интегрированной системы для научных библиотек, архивов и музеев // Информационные ресурсы России. – 2012. – № 1. – С. 2–7
3. Горный, Е. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы / Е. Горный, К. Вигурский // Интернет и российское

- общество / Под ред. И. Семенова; Моск. Центр Карнеги. – М.: Гендальф, 2002. С.158–188.
4. <http://www.w3.org/standards/semanticweb/>
  5. C. Bizer, T. Heath, and T. Berners-Lee. Linked data – the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1—22, 2009.
  6. <http://www.w3.org/RDF/>
  7. T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993
  8. Lihua Zhao. Ontology Integration for the Linked Open Data // 2013 <http://ri-www.nii.ac.jp/~lihua/PhD.Thesis.LihuaZhao.pdf>
  9. <http://e-heritage.ru/index.html>
  10. Candela L., Castelli D., Dobreva M., Ferro N., Ioannidis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D. The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST–2002 2.3.1.12. Technology–enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007. [http://www.delos.info/files/pdf/ReferenceModel/DELOS\\_DLReferenceModel\\_0.98.pdf](http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf)
  11. Erhard Rahm, Hong Hai Do. Data Cleaning: Problems and Current Approaches. 2000
  12. Ranjit Singh, Dr. Kawaljeet Singh A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing // IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 2, May 2010
  13. Scannapieco M., Missier P., Batini C. Data Quality at a Glance // Datenbank–Spektrum. – 2005. – Т. 14. – С. 6–14.
  14. Sebastian Ryszard Kruk • Bill McDaniel. Semantic Digital Libraries // 2009 – Springer–Verlag Berlin Heidelberg

# ИНФОРМАЦИОННЫЕ СИСТЕМЫ И СЕРВИСЫ КОМПЛЕКСНОЙ ПОДДЕРЖКИ ПЕРИОДИЧЕСКИХ НАУЧНЫХ ИЗДАНИЙ

Д. Ю. Ахметов<sup>1</sup>, А. М. Елизаров<sup>1</sup>, Е. К. Липачёв<sup>1</sup>

*1 Институт математики и механики им. Н. И. Лобачевского  
Казанского (Приволжского) федерального университета*

**Введение.** Сегодня общепризнано, что после получения нового научного знания и даже отдельного конкретного научного результата время, необходимое для их доведения до заинтересованного пользователя (ученого-теоретика, инженера-практика, менеджера, преподавателя, студента и других) должно быть максимально коротким. Именно по этой причине даже традиционные задачи, связанные с подготовкой научной работы и последующими процессами публикации статьи в научном журнале или материалах научной конференции и информирования заинтересованных читателей, решаются в настоящее время на новых организационном и информационном уровнях и базируются на широком использовании интернета и информационно-коммуникационных технологий (ИКТ). Кроме того, существенно изменилась инфраструктура современных научных изданий – речь уже идет не столько о формах и средствах использования ИКТ, сколько о создании программной платформы, реализующей развитую систему сервисов для работы с электронным контентом. При этом стираются различия между электронным и печатным изданиями, являющимися лишь различными формами представления соответствующей информации; без информационных сервисов и специализированных программ становится невозможным подготовить научную работу к публикации, а без сетевых коммуникаций – опубликовать ее. Учитывая, что в любом случае речь идет об использовании электронного контента, можно говорить об электронных научных изданиях, отдельно не выделяя печатные их формы, а электронные научные публикации можно рассматривать как разновидность электронных документов, особенности которых отражаются в их жизненном цикле.

Таким образом, современный электронный научный журнал должен базироваться на программной платформе, реализующей развитую систему сервисов для работы с электронным контентом. Отметим, что базовые понятия «электронный документ» и «жизненный цикл электронного документа», свойства электронных научных документов и их преимущества перед традиционными бумажными носителями обсуждены в [1].

В настоящей работе обсуждаются сервисы поддержки жизненного цикла научной публикации, определяющие функциональность современных программных платформ управления электронными научными журналами.



Особое внимание уделено сервису поддержки научного рецензирования – описан разработанный алгоритм автоматизации экспертной оценки, указаны изменения как в программном коде открытой системы Open Journal Systems, являющейся ядром программной платформы, так и в используемой базе данных.

**Жизненный цикл научной публикации и сервисы его поддержки.** Как известно, в сфере ИКТ для обозначения последовательности процессов от проектирования, разработки и сопровождения информационной системы до прекращения её эксплуатации используется термин «жизненный цикл». Каждый документ, в том числе электронный, за время своего существования также проходит определенные стадии, называемые «жизненным циклом документа». Обычно выделяют следующие его этапы: создание, описание, утверждение, хранение, поиск, извлечение и доставка. Жизненный цикл электронной научной публикации, в отличие от «бумажной», предполагает дальнейшее улучшение её качества и расширение свойств. Отметим в этой связи, например, новую технологию в области электронных научных публикаций, получившую название «живых документов» (см., например, [2, 3]). Как отмечено в [4], онлайн-представление становится центральным звеном публикационной деятельности ученого, и с него начинается жизненный цикл публикации.

Какими же должны быть сервисы поддержки жизненного цикла научной публикации? Анализ функциональности, реализуемой многочисленными современными программными платформами управления электронными научными журналами различных научных издательств, показал, что эта функциональность представлена рядом основных и дополнительных (опциональных) сервисов. В целом все их можно разделить на следующие группы: *сервисы подготовки научной публикации* – *сервисы для автора* (подготовки научной статьи в соответствии с редакционными требованиями, а также взаимодействия с редколлегией журнала), *сервисы поддержки научного рецензирования* (рецензента, редактора), *сервисы поддержки редакционных процессов* и *сервисы для читателя*. Все эти сервисы непосредственно связаны с процессами, реализуемыми с участием авторов и (или) редколлегии (редакции) журнала.

К основным процессам, связанным с деятельностью авторов и редколлегии (редакции) электронного научного журнала на начальном этапе, традиционно относят представление научной публикации, ее рецензирование и редактирование.

На этапе представления научной публикации важна автоматическая проверка статьи на соответствие стиливым правилам журнала (стилевая валидация), следовательно, нужна организация соответствующего автоматизированного сервиса подготовки научной публикации. В частности, в физико-математических статьях, набранных в T<sub>E</sub>X-нотации, должны использоваться стилевые файлы, специально подготовленные редакцией, и

соответствующие макрокоманды. Для физико-математических журналов необходима также программная поддержка процесса обработки электронных документов, содержащих формулы. Для работы с электронными документами, подготовленными в T<sub>E</sub>X-нотации, а это основной формат физико-математических журналов, разработан специализированный модуль стилевого оформления публикации по математике, размещенный на платформе science.tatarstan.ru (см. [5]). Соответствующий сервис выполнен в виде конструктора, предлагающего автору заполнить ряд полей, по которым автоматически формируется T<sub>E</sub>X-код статьи, выполненный по правилам журнала.

Другим сервисом для авторов является конструктор создания библиографических списков статей, учитывающий правила научного журнала и принятые сокращения названий журналов.

Еще одним сервисом для авторов, абсолютно необходимым при размещении информации о статье в цитатных базах данных (Scopus, Web of Science), является транслитерация библиографии публикации, т. е. процедура точной передачи в библиографическом списке статьи букв алфавита одного языка (и, соответственно, отдельных слов и текстов) средствами другой системы (алфавита другого языка), в частности, русских слов латиницей.

К сервисам поддержки научного рецензирования относится, прежде всего, автоматизация выбора рецензентов и самого процесса рецензирования. Отметим, что один из вариантов такой автоматизации реализован на платформе science.tatarstan.ru (см. [6]).

К редакционным процессам относят классификацию, аннотирование, выделение метаданных, публикацию, объединение в коллекцию, организацию поиска и навигации, долгосрочное хранение, конвертирование в различные форматы и распространение, контроль доступа, подписку, рассылку уведомлений, анализ статистики использования (включая формирование наукометрических данных). Сервисы, обеспечивающие автоматизацию перечисленных редакционных процессов, должны обеспечиваться любой информационной системой управления электронным научным журналом, их реализация на портале журнала, безусловно, необходима, однако набор таких сервисов должен постоянно расширяться в целях устойчивого развития издания. В частности, дополнительными сервисами поддержки редакционных процессов являются:

- проверка загружаемых текстов на плагиат;
- автоматизированная разметка статей, например, в математических электронных коллекциях (см. [7]);
- поиск и сбор OAI-метаданных (например, с помощью Open Harvester Systems);
- информетрический анализ (например, поддержка сервиса Article-Level Metrics (<http://article-level-metrics.plos.org>) в системе OJS);
- поддержка научных конференций (например, системы [www.easychair.org](http://www.easychair.org))

и Open Conference System);

- интеграция новостных лент научных журналов.

К набору сервисов для читателя можно отнести:

- оплату услуг (например, возможность работы с электронным кошельком PayPal ([www.paypal.com](http://www.paypal.com)));

- онлайн-общение (вебинары; видеоконференции);

- учет специфики обрабатываемых информационных ресурсов, например, расширенный поиск по фрагментам формул в математических коллекциях [7, 8], а также поиск на основе платформы семантической публикации математических документов [8, 9].

Итак, современные информационные системы управления электронными научными публикациями не должны ограничиваться сервисами удаленного представления статей в научный журнал и их дальнейшей обработки для окончательной публикации – в первую очередь необходимы доступ к сформированному контенту и расширенный поиск (по автору, названию статьи, ключевым словам и др.) в соответствующих электронных коллекциях. Тогда в полном объеме будут реализованы функциональные возможности, присущие электронным библиотекам.

**Программные платформы управления бизнес-процессами научного журнала.** Все названные выше функции и сервисы целесообразно обеспечить в рамках единой программной платформы. Внедрение такой платформы позволит, прежде всего, автоматизировать наиболее трудоемкие рабочие процессы, а порталное решение даст возможность интегрировать журнал в глобальное информационное научное пространство. Вместе с тем, для развития каждого электронного научного журнала необходимо постоянное расширение функциональности его базовой информационной системы.

В настоящее время успешное решение проблемы обеспечения оперативной доступности результатов научных исследований становится все теснее связанным с внедрением информационных издательских систем, разрабатываемых как в научных сообществах, так и в издательской отрасли в целом. Разработано и функционирует большое число информационных систем поддержки издательских процессов, которые делятся на системы с открытым и закрытым исходным кодом [10]. Последние принадлежат, как правило, ведущим мировым издательствам, обрабатывают по единым стандартам, с использованием единых технологий и единообразных требований тысячи журналов и не доступны небольшим издательствам (разве что на условиях аутсорсинга и, естественно, не бесплатно), выпускающим до нескольких десятков научных изданий. В такой ситуации небольшие издательства (в частности, издательства многих университетов и академических НИИ) при переходе в своей работе на современные ИКТ либо самостоятельно разрабатывают соответствующие информационные системы (которые, как правило, малоэффективны, не конкурентоспособны и, самое главное, не тиражируемы), либо используют системы с открытым кодом, дорабатывая их с

учетом собственных потребностей.

Среди систем управления электронными научными журналами с открытым кодом несомненным лидером стала программная платформа Open Journal Systems (OJS): согласно статистике OJS, в 2013 году система использовалась более чем в 6800 активных журналах, и это количество постоянно росло с момента начала ее внедрения (<https://pkp.sfu.ca/ojs/ojs-usage/ojs-map/>). Отметим, что преимущества использования систем с открытым кодом и выбора системы OJS подробно описаны в [10]. Там же предложена трехуровневая архитектура платформы управления электронными научными журналами, использующая в качестве ядра систему OJS.

**Технология расширения функционала OJS.** Система OJS имеет модульную архитектуру, что позволяет разрабатывать собственные классы и модули, расширяя ее функциональные возможности. OJS основана на паттерне Model-View-Controller (MVC), как следствие, хранилище данных, пользовательские интерфейсы и управляющие функции разделены на разные уровни взаимодействия. Система OJS платформонезависима и может быть установлена как под ОС Windows, так и на Unix-ориентированных операционных системах. В OJS используются свободно распространяемые язык программирования PHP, веб-сервер Apache, а также СУБД (MySQL, PostgreSQL).

Расширить функционал системы можно с помощью модулей, имеющих специальный формат (см., например, [11]). Большое количество инструментов, расширяющих возможности пользовательского интерфейса системы, а также обеспечивающих поддержку различных форматов данных, содержится в галерее модулей (плагинов) OJS (<http://pkp.sfu.ca/support/forum/viewforum.php?f=28>).

При самостоятельной разработке плагина для OJS необходимо вначале определить «тип» плагина, например, плагины, определяющие способы авторизации, относятся к типу «плагин авторизации», при разработке файлы плагина должны быть расположены в директории `\plugins\auth`. Плагины, не попадающие ни в один из типов, определяются как «общие» и должны быть расположены в директории `\plugins\generic`.

Как показал наш опыт, при разработке таких плагинов целесообразно следовать следующим правилам:

- все файлы создаваемого плагина должны быть размещены в отдельной папке;
- служебную информацию о плагине (дата создания, версия, тип, имя и др.) сохранять в xml-формате (например, в файле `version.xml`);
- код плагина подготовить на языке PHP и сохранить в файле `myPlugin.inc.php`, где `myPlugin` – название плагина;
- создать файл `index.php` для вызова плагина `myPlugin.inc.php`.

**Сервис поддержки научного рецензирования.** Важная отличительная особенность работы редколлегии научного журнала – наличие системы

независимого рецензирования и, как следствие, группы рецензентов-экспертов в предметных областях, определенных тематикой журнала. Процесс научного рецензирования – наиболее длительный этап в цепочке прохождения статьи в редакции научного журнала (от рукописи до публикации). В печатных научных журналах этот процесс мог затягиваться на годы. Критическими здесь, прежде всего, являются трудности в подборе рецензентов. Автоматизация этого процесса – выбор рецензента из базы экспертов с учетом не только компетенции эксперта, но и объема рецензирования и количества рецензируемых работ, – существенно сокращает сроки рецензирования. Отметим, что система автоматического назначения рецензента успешно использовалась в журнале *Lobachevskii Journal of Mathematics* [12].

Разработанный нами алгоритм автоматизации экспертной оценки (рецензирования) статьи, представленной в журнал, основан на использовании ключевых слов, которые в соответствии с установленными правилами должны быть указаны в статье, а также представлены автором при регистрации статьи в информационной системе научного журнала. С помощью ключевых слов, а также классификационных признаков (УДК, ББК, *Mathematics Subject Classification*) можно в автоматическом режиме достаточно точно определить область исследований, к которой относится представленная статья, и сопоставить эти данные с информацией об экспертах, содержащейся в базе данных информационной системы журнала [13, 14].

Решить задачу установления соответствия между множеством ключевых слов рассматриваемых научных работ и множеством, характеризующим специализацию рецензентов, можно различными способами. Наиболее простым (в плане технической реализации) является выделение одного из множеств в качестве базового – таковым целесообразно взять множество специализаций рецензентов и затем для каждой статьи определить несколько ключевых слов из этого множества. Форму отзыва рецензента целесообразно расширить за счет возможности добавления комментариев для выяснения мотивации проведенной им оценки.

Процесс экспертной оценки организован по стандартному принципу «принять в печать» – «вернуть с замечаниями на доработку» – «отклонить». В автоматическом режиме осуществляются контроль сроков рецензирования, а также рассылка уведомлений авторам и рецензентам. При выборе рецензентов система автоматически учитывает загруженность эксперта за счет ведения «шкалы загруженности» и истории рецензирования. Исключается возможность саморецензирования – эксперт, представивший статью в журнал в качестве одного из авторов, не сможет попасть в список возможных ее рецензентов.

Основные этапы алгоритма автоматического выбора экспертов научной статьи заключаются в следующем:

- формирование входных параметров – в качестве входного аргумента используется идентификатор научной работы (или первичный ключ сущности «научная работа» на языке баз данных);

- создание общего списка рецензентов – выборка осуществляется из базы данных рецензентов с учетом их специализации, загруженности, а также условия «рецензент не может быть автором»;
- отбор рецензентов по критерию наилучшего соответствия (большее число совпадений по ключевым словам, минимальные уровни загруженности, учет рейтинга рецензента и истории рецензирования);
- рассылка уведомлений рецензентам, авторам и редакторам.

Реализация описанного алгоритма потребовала изменений как в программном коде OJS, так и в используемой базе данных. Ограничения на возможность включения плагинов в систему OJS, связанные с использованием языка программирования PHP, потребовали изменения файла `peerReview.tpl`, а именно, к пункту меню «Выбрать рецензента» добавлен новый пункт меню «Automatic selection of the reviewer», перенаправляющий систему к плагину автоматизации. Параметры работы скрипта передаются методом GET, поэтому идентификатор статьи включается в адресную строку с указанием абсолютного пути к файлу скрипта:

```
<form>
  {assign var="idij" value=$submission->getId()}
  <a href='http://.../plugins/generic/autoReviewer/myPlugin.inc.php
    ?iu={$idij}'>Automatic selection of the reviewer</a>
</form>
```

После активации плагина считывается значение параметров из адресной строки и формируется список первичных ключей таблицы «roles», значения «user\_id» которых равно 4096 (т.к. данные обо всех пользователях OJS хранятся в одной таблице базы данных, каждая группа идентифицируется определенным натуральным числом, роль рецензента определяется числом 4096). Далее создаются массивы, состоящие из идентификаторов рецензентов и уровня их загруженности, производятся сортировка и отбор рецензентов. Формируются переменные «\$newdate» и «\$newdate2» для вычисления текущего времени и времени, к которому истекает срок подачи рецензии (по умолчанию OJS для составления рецензии выделяет 4 недели):

```
$newdate=date("Y-m-d H:i:s");
$newdate2=date ('Y-m-d 00:00:00', strtotime ('+4 weeks')).
```

В базе данных устанавливается связь между статьей и рецензентами, для этого в таблице «review\_assignments» создаются записи, количество которых отражает число отобранных рецензентов. В поля `submission_id` (идентификатор статьи), `reviewer_id` (идентификатор рецензента), `date_assigned` (текущее время) и `date_due` (конечная дата подачи рецензии) записываются значения '`$article`', '`$massiv_index[$index]`', `op'``$newdate`', '`$newdate2`'. После окончания всех операций с данными при помощи скрипта JavaScript организуется возврат в систему:

```
echo (" <script> location.href = 'http://.../editor/submissionReview/$article' </script>").
```

Вместе с тем, автоматический подбор рецензентов – только часть процесса автоматизации, необходимы также вспомогательные инструменты рецензирования, обеспечивающие загрузку наукометрических данных, поиск по статьям с близкой тематикой, сравнение полученных результатов с опубликованными ранее, в частности, на наличие заимствований и полноту ссылок.

Если научная статья прошла процесс рецензирования и получила положительные отзывы, то начинается редакционный процесс. На этом этапе происходят литературное редактирование статьи и ее перевод в форматы, поддерживаемые информационной системой журнала (pdf, html и др.). После этого рассматриваемая статья проходит техническое редактирование и при обнаружении ошибок вновь подвергается редакционному процессу. На данном этапе определяем также номер выпуска электронного издания, в котором статья будет опубликована. Процесс публикации непосредственно включает в себя компиляцию подготовленного электронного издания, устранение недочетов и технических неточностей.

**Особенности обработки математических текстов.** Большая часть электронных математических коллекций состоит из неструктурированных текстов в различных форматах. Ведущими мировыми издательствами физико-математической литературы, в том числе российскими, принята  $T_E X$ -нотация. Теговая система  $T_E X$ -документа позволяет придать ему структурную выразительность, недоступную другим системам нотации,  $T_E X$ -нотация обладает практически всеми необходимыми компонентами для создания семантической разметки, что особенно важно при создании процедур автоматической обработки математических документов электронной коллекции и использовании программ конвертации. Для работы с  $T_E X$ -документами функционал системы OJS необходимо расширить путем подключения специальных модулей, например, с помощью плагина, предложенного в [6].

Успешными примерами создания программной платформы управления математическими научными журналами служат широко известный портал Math-Net.Ru (<http://www.mathnet.ru/>) [15-17] и издательская платформа журнала Lobachevskii Journal of Mathematics [6, 12–14].

Работа выполнена при финансовой поддержке РФФИ (проекты 15-07-08522, 15-47-02472) и РГНФ (проект 14-03-12004).

## ЛИТЕРАТУРА

1. Елизаров А.М., Зуев Д.С., Липачёв Е.К. Управление жизненным циклом электронных публикаций в информационной системе научного журнала // Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии. – 2014. – № 4. – С. 81–88.
2. Горбунов-Посадов М.М. Живая публикация // Открытые системы. – 2011. – № 4. – С. 48-49. Обновляемая («живая») публикация: URL: <http://keldysh.ru/gorbunov/live.htm> .

3. Паринов С.И., Когаловский М.Р. «Живые» документы в электронных библиотеках // Прикладная информатика. – 2009. – №6 (24). Авторская версия: <http://socionet.ru/publication.xml?h=repec:rus:isyigw:article-215>.
4. Горбунов-Посадов М.М. Интернет-активность как обязанность ученого // Информационные технологии и вычислительные системы. – 2007. – № 3. – С. 88-93. – URL: <http://keldysh.ru/gorbunov/duty.htm>.
5. Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К. Автоматизация процесса первичной обработки математической статьи в информационной системе электронного научного журнала // Тр. Математического центра имени Н.И. Лобачевского. Материалы Двенадцатой молодежной науч. шк.-конф. «Лобачевские чтения – 2013». – Казань: Изд-во Казан. матем. об-ва, 2013. – Т. 47. – С. 6-10.
6. Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К. Система автоматизации редакционных процессов на платформе электронных научных журналов // Учёные записки Института социальных и гуманитарных знаний. – 2014. – №1 (12), Ч. 2. – С. 228-233.
7. Biryal'tsev E., Elizarov A., Zhil'tsov N., Lipachev E., Nevzorova O., Solov'ev V. Methods for analyzing semantic data of electronic collections in mathematics // Automatic Documentation and Mathematical Linguistics. – 2014. – V. 48, No 2. – P. 81-85.
8. Елизаров А.М., Липачев Е.К., Малахальцев М.А. Веб-технологии для математика. Основы MathML. – М.: Физматлит, 2010. – 192 с.
9. Елизаров А.М., Липачёв Е.К., Невзорова О.А., Соловьев В.Д. Методы и средства семантического структурирования электронных математических документов // Докл. РАН. – 2014. – Т. 457, № 6. – С. 642-645.
10. Елизаров А.М., Зуев Д.С., Липачёв Е.К. Информационные системы управления электронными научными журналами // Научно-техническая информация. Серия 1: Организация и методика информационной работы. – 2014. – № 3. – С. 31-38.
11. Willinsky J., Stranack K., Smecher A., MacGregor J. Open Journal Systems: a complete guide to online publishing. – Simon Fraser University Library, 2010. – 273 p.
12. Елизаров А.М., Липачёв Е.К., Хохлов Ю.Е. Семантические методы структурирования математического контента, обеспечивающие расширенную поисковую функциональность // Информационное общество. – 2013. – № 1-2. – С. 83-92.
13. Ахметов Д.Ю. Управление жизненным циклом электронной научной публикации // Тр. XV Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции, RCDL-2013». Ярославль: Изд-во ЯрГУ им. П.И. Демидова, 2013. – С. 407-408.
14. Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К. Модель сервисов электронного математического журнала и ее облачная реализация на платформе Open Journal Systems // Труды Российской школы «Математическое и



компьютерное моделирование фундаментальных объектов и явлений» и Международного научного семинара «Нелинейные поля в теории гравитации и космологии». – Казань: Отечество, 2013. – С. 86-92.

15. Жижченко А.Б., Изаак А.Д. Информационная система Math-Net.Ru. Применение современных технологий в научной работе математика // Успехи матем. наук. – 2007. – Т. 62, Вып. 5 (377). – С. 107-132.

16. Жижченко А.Б., Изаак А.Д. Информационная система Math-Net.Ru. Современное состояние и перспективы развития. Импакт-факторы российских математических журналов // Успехи матем. наук. – 2009. – Т. 64, Вып. 4 (388). – С. 195-204.

17. Chebukov D.E., Izaak A.D., Misyurina O.G., Pupyrev Y.A., Zhizhchenko A.B. Math-Net.Ru as a digital archive of the russian mathematical knowledge from the XIX century to today // Lecture Notes in Computer Science, V. 7961, ed. J. Carette et al., 2013. – P. 344-348, arXiv: 1305.5655

# ИСПОЛЬЗОВАНИЕ ИНТЕРНЕТА ДЛЯ ОБУЧЕНИЯ ПАРАЛЛЕЛЬНОМУ ПРОГРАММИРОВАНИЮ

**В.А. Бахтин, О.Ф. Жукова, Н.А. Катаев, А.С. Колганов, В.А. Крюков,  
Н.В. Поддерюгина, М.Н. Притула, О.А. Савицкая, А.А. Смирнов**

*Институт прикладной математики им. М.В. Келдыша РАН*

Количество пользователей, ежедневно входящих в сеть Интернет, активно растет. По данным Всероссийского центра изучения общественного мнения (ВЦИОМ) доля интернет-пользователей в 2015 году составляет 69%, а доля тех, кто выходит в Сеть ежедневно, составляет половину россиян (52%). Доля интернет-пользователей среди молодежи (18-24 летних) составляет 96%. Самыми популярными поводами обращения к Сети являются медиа-ресурсы (56%), поиск новостей и информации (55%), электронная почта (53%), работа и учеба (50%). Опрос проведен 21-22 марта 2015 года, опрошено 1600 человек в 130 населенных пунктах в 46 областях, краях и республиках России [1]. Результаты опроса показывают, что, поскольку все большее количество молодых людей пользуются Интернетом, Всемирная сеть становится важным ресурсом для использования в сфере образования. Это ставит перед преподавателями ряд вопросов. Нужно, чтобы информационная среда была не просто доступна студентам, но и удовлетворяла следующим требованиям. Прежде всего, она должна содержать достоверную информацию, включать общепринятую терминологию, подачу информации следует осуществлять от простого к сложному, облегчая тем самым усвоение ее студентами. Это должно сопровождаться укреплением связи студент-преподаватель, в том числе и для усиления контроля преподавателя за образовательным процессом.

Активное использование подхода Веб 2.0 к организации Интернет-ресурсов ведет к тому, что бесконтрольный поиск информации приводит студентов на ресурсы, достоверность которых может вызывать сомнения. Следует обратить внимание, что в интернет-энциклопедии Википедия на странице, озаглавленной «Отказ от ответственности» [2], крупными буквами указано «Википедия не гарантирует истинности». Это связано с тем, что приведенные в ней данные «могли быть только что изменены, испорчены, либо написаны тем, чье мнение отличается от общепринятого в области знаний, интересующей вас».

Кроме того, поиск нужной информации может сам по себе вызывать значительные трудности, связанные с колоссальным объемом данных, составляющих интернет-пространство [3]. Огромную роль играет формулировка запроса, правильно указанные ключевые слова. Если поиск проводится в рамках специализированной области, такой как параллельное программирование, важную роль играет знание общепринятой терминологии, в

том числе на иностранных языках. В качестве примера рассмотрим исследования, проводимые в области статического анализа программ. Результаты анализа применяются в автоматически распараллеливающих компиляторах, в системах автоматизированного распараллеливания, таких как САПФОР [4]. Для эффективного распараллеливания циклов последовательной программы необходимо указать, какие переменные могут быть объявлены приватными. В терминологии OpenMP [5] такие переменные должны быть перечислены в директивах `private`. Поиск по запросу `private variables` в поисковой системе Google не дает нужных результатов, так как аналогичная терминология используется при ограничении области видимости во многих языках программирования, не связанных с параллелизмом. Указав вместо `private` – `privatizable`, что точнее соответствует цели поиска (переменные, которые могут быть использованы как приватные) и, добавив ключевое слово `analysis`, можно найти статьи, посвященные соответствующим алгоритмам, применяемым при статическом анализе программ.

Справиться с перечисленными трудностями можно, подготовив ссылки на электронные ресурсы с проверенной информацией по изучаемой проблеме, подготовив ресурсы, описывающие задания и контролирующие их выполнение, наладив электронную переписку, в том числе организовав списки рассылки. Возможно привлечение различных средств связи через Интернет в режиме реального времени (например, Skype).

При обучении параллельному программированию особое внимание следует уделить доступу к вычислительным ресурсам. Суперкомпьютерные системы доступны только удаленно и, как правило, сильно загружены. Это не позволяет оперативно выполнять большое количество практических заданий, необходимых в процессе обучения. Наряду с доступом к мощным вычислительным комплексам, доступ к которым необходим для полного понимания изучаемой области, необходимо организовывать доступ к небольшим рабочим станциям для практического совершенствования навыков параллельного программирования.

Спецсеминар «Операционные системы и языки программирования распределенных вычислительных систем» кафедры Системного программирования (СП) факультета Вычислительной математики и кибернетики (ВМиК) Московского государственного университета им М.В. Ломоносова (МГУ) [6] проводит подготовку в области суперкомпьютерных технологий. Работа спецсеминара связана с исследованиями, проводимыми в направлениях разработки высокоуровневых языков параллельного программирования и создания систем автоматизированного распараллеливания. Студенты получают теоретическую подготовку и приобретают опыт разработки параллельных программ для современных вычислительных кластеров. Исследовательская деятельность студентов и аспирантов направлена на разработку инструментов параллельного программирования. При их активном участии развиваются системы DVM [7, 8]

и САПФОР [4], разработанные в Институте прикладной математики им. М.В. Келдыша РАН.

Обучение программированию невозможно без выполнения практических заданий, и параллельное программирование – не исключение. Система контроля выполнения практических заданий организована на базе веб-приложения для управления проектами Redmine [9]. Прямое назначение данного инструмента – управление проектами, задачами, отслеживание ошибок. Инструмент используется при разработке систем DVM и САПФОР для распределения задач между разработчиками, для отслеживания статуса задач и ошибок, обнаруживаемых в системе, для расставления приоритетов задач. Каждой из разрабатываемых систем в Redmine соответствует отдельный проект.

При обучении параллельному программированию создается учебный проект, в котором для студентов распределяются задания. Таким образом, студенты одновременно с изучением параллельного программирования приобретают навыки по управлению проектами, и для привлечения их к участию в разработке систем DVM или САПФОР не требуется дополнительного изучения данного инструмента.

Каждый студент, участвующий в спецсеминаре, регистрируется в системе Redmine, при этом создается аккаунт, связанный с почтовым адресом студента. В процессе работы над проектом на указанный адрес приходит информация о статусе выполняемых студентом задач, отправляются оповещения о появлении новых задач, о том, считается ли задача сданной. Система Redmine позволяет отвести каждому участнику проекта определенную роль, необходимым образом ограничив доступ к функциям системы. Каждый студент получает роль Student. Затем ему будет поставлена отдельная задача, которую он должен выполнить. В каждом проекте создается страница Wiki с описанием общей части поставленных задач (см. Рис. 1).

Преподаватели спецсеминара также зарегистрированы в системе и участвуют в учебном проекте в роли Manager. С каждой задачей связан аккаунт одного или нескольких преподавателей, отслеживающих ее выполнение. При изменении состояния задачи на электронные адреса преподавателей приходят соответствующие уведомления.

Выполнение задачи происходит в несколько этапов, сложность каждого этапа постепенно возрастает. Каждый этап оформляется в системе как подзадача. На каждом этапе студент последовательно знакомится с различными технологиями параллельного программирования (OpenMP, OpenACC [10], DVMH), а также изучает средства профилирования программ, такие как Intel VTune Amplifier [11].

Home My page Projects Administration Help

Logged in as: N\_Kataev My account Sign out

Practical training 2014

Search:  Practical training 2014

Overview Activity Issues New issue Gantt Calendar Wiki Files Settings

### Задание 2014 г.

В 2014 году каждому студенту выдаются в качестве исходных две Фортран-программы – последовательная версия одного из тестов NAS NPВ 3.3.1 и ее параллельная версия в модели OpenMP. Требуется распараллелить последовательную программу в моделях OpenACC и DVMН. Это задание разбивается на следующие этапы, каждому из которых в итоговом отчете соответствует раздел.

- 1. Получение на целевой ЭВМ времен выполнения**
- 2. Профилирование выполнения последовательной программы**
- 3. Профилирование выполнения OpenMP-версии на 1, 2, 4, 8, 12 нитях**
- 4. Распараллеливание программы на OpenACC и получение на целевой ЭВМ времен выполнения на одном ГПУ**
- 5. Распараллеливание программы на DVM и получение на целевой ЭВМ времен выполнения**
- 6. Распараллеливание программы на DVMН и получение на целевой ЭВМ времен выполнения**

Доступ к целевым ЭВМ осуществляется по SSH

Получение на целевой ЭВМ (например, K-100, Ломоносов или др.) для нескольких классов начальных данных (классы А, С) следующих времен выполнения:

- время выполнения последовательной версии;
- время последовательного выполнения OpenMP-версии;
- время выполнения OpenMP-версии на 1, 2, 4, 8, 12 нитях одного узла.

**2. Профилирование выполнения последовательной программы**

- получение информации с помощью инструментов Интела;
- получение информации о временах выполнения всех циклов с помощью инструментов DVM (задание опции -e4, расширение fdv, инструкция PA);
- определение самых времянких циклов;
- оценка накладных расходов профилирования.

Профилирование последовательной и OpenMP-программы может осуществляться при помощи:

- Intel VTune Amplifier XE 2013
- Intel Advisor XE

Данные инструменты входят в состав Intel Cluster Studio XE 2013, Intel Parallel Studio XE или Intel Fortran Studio XE. Пробные версии для Windows (на 30 дней) или версию для некоммерческого использования для Linux можно скачать с сайта <http://software.intel.com/ru-ru/>

Для профилирования OpenMP-программы на K-100 можно использовать опцию -openmp-profile. В результате запуска программы, скомпилированной с опцией -openmp-profile, на выполнение создается текстовый файл guide.gvs - программа разбивается на последовательные и параллельные участки, и для каждого такого участка собираются характеристики эффективности.

Также для профилирования может быть полезен AMD CodeXL: <http://developer.amd.com/tools-and-sdks/heterogeneous-computing/codex/>

- 3. Профилирование выполнения OpenMP-версии на 1, 2, 4, 8, 12 нитях**
- 4. Распараллеливание программы на OpenACC и получение на целевой ЭВМ времен выполнения на одном ГПУ**
- 5. Распараллеливание программы на DVM и получение на целевой ЭВМ времен выполнения**
- 6. Распараллеливание программы на DVMН и получение на целевой ЭВМ времен выполнения**

Распараллеливание программы на DVM и получение на целевой ЭВМ времен выполнения на 1, 2, 4, 8, 12 процессах одного узла, а также времени выполнения этой DVM-программы в последовательном режиме.

Распараллеливание программы на DVMН и получение на целевой ЭВМ времен выполнения на 1,2,3 ГПУ.

Рис. 1. Пример задания, выполняемого студентами в 2014 году

На Рис. 2 приведен пример отображения состояния задачи по распараллеливанию теста из пакета NAS NPВ [12]. Основная задача состоит из шести подзадач, соответствующих определенным этапам обучения. Для студенческих задач допускается четыре стадии выполнения. New означает, что задача только создана и студент еще не приступил к ее выполнению. In Progress устанавливается студентом для решаемых в данный момент задач. Resolved устанавливается студентом, когда он считает, что задача выполнена. Статус Closed может быть установлен только преподавателем после проверки корректности выполнения задачи. В процессе выполнения студент указывает время, которое он затратил на каждый этап задания и, если задача еще не завершена, процент от ее полной готовности. Когда задача считается законченной, к ее описанию добавляется отчет о выполнении (см. Рис. 3).

#	Tracker	Status	Priority	Subject	Updated	% Done
232	Feature	In Progress	Normal	Распараллеливание теста NAS MG на OpenACC и DVMH	22.12.2014 22:10	<div style="width: 100%;"></div>
233	Support	Closed	Normal	MG Получение на целевой ЭВМ времен выполнения	25.05.2015 23:22	<div style="width: 100%;"></div>
234	Support	Closed	Normal	MG Профилирование выполнения последовательной программы	25.05.2015 23:23	<div style="width: 100%;"></div>
235	Support	Resolved	Normal	MG Профилирование выполнения OpenMP-версии на 1, 2, 4, 8, 12 нитях	28.10.2014 09:07	<div style="width: 100%;"></div>
236	Feature	Resolved	Normal	MG Распараллеливание программы на OpenACC и получение на целевой ЭВМ времен выполнения на одном ГПУ	02.12.2014 16:52	<div style="width: 100%;"></div>
237	Feature	In Progress	Normal	MG Распараллеливание программы на DVM и получение на целевой ЭВМ времен выполнения	22.12.2014 22:10	<div style="width: 100%;"></div>
238	Feature	New	Normal	MG Распараллеливание программы на DVMH и получение на целевой ЭВМ времен выполнения	10.09.2014 12:11	<div style="width: 100%;"></div>

Рис. 2. Пример отображения состояния выполняемой задачи

**Support #233**

Feature #232: Распараллеливание теста NAS MG на OpenACC и DVMH  
**MG Получение на целевой ЭВМ времен выполнения**

Added by Nikita Kataev 9 months ago. Updated 14 minutes ago.

**Status:** Closed  
**Priority:** Normal  
**Assignee:** -  
**Category:** -  
**Target version:** -

**Start date:** 01.10.2014  
**Due date:** 07.10.2014  
**% Done:** 100%  
**Estimated time:** 16.00 hours

**Description**

Получение на целевой ЭВМ (например, K-100, Ломоносов или др.) для нескольких классов начальных данных (классы A, C) следующих времен выполнения:

- время выполнения последовательной версии;
- время последовательного выполнения OpenMP-версии;
- время выполнения OpenMP-версии на 1, 2, 4, 8, 12 нитях одного узла.

**Subtasks**

**Related issues**

**History**

Updated by Nikita Kataev 9 months ago #1

- copied from Support #219: VT Получение на целевой ЭВМ времен выполнения added

Updated by Nikita Kataev 9 months ago #2

- copied from deleted (Support #219: VT Получение на целевой ЭВМ времен выполнения)

Updated by Nikita Kataev 9 months ago #3

- Subject changed from VT Получение на целевой ЭВМ времен выполнения to MG Получение на целевой ЭВМ времен выполнения
- Assignee changed from [redacted] to [redacted]

Updated by [redacted] 8 months ago #4

- Status changed from New to Resolved
- % Done changed from 0 to 100

Получены времена выполнения теста MG на K-100:

CLASS	Последовательный вариант	Последовательный OpenMP вариант	OpenMP 1 процесс	OpenMP 2 процесса	OpenMP 4 процесса	OpenMP 8 процессов	OpenMP 12 процессов
A	1.41s	1.41s	1.56s	0.83s	0.42s	0.31s	0.33s
C	55.71s	55.38s	60.48s	31.01s	18.42s	13.87s	13.38s

Рис. 3. Подробное описание задачи

На определенном этапе обучения студенты приступают к изучению DVM-системы для разработки параллельной версии программы для кластера с графическими ускорителями. Распараллеливание выполняется с помощью DVMH-расширения стандартных языков параллельного программирования Fortran или C. Знакомство с DVM-системой начинается с обращения к сайту данного проекта [13], созданного авторами статьи.

Цель создания сайта DVM-системы – предоставить пользователям разного уровня подготовки, в том числе студентам, изучающим параллельное программирование, актуальную информацию о возможностях системы, заинтересовать и привлечь новых пользователей, обеспечить удобный доступ к системе и оперативную обратную связь с пользователями. Доступны русскоязычная и англоязычная версии сайта. Выделяется пять основных разделов: главная страница, документация, получение доступа к системе, публикации и доклады, презентующие различные особенности системы, контактная информация.

Попадая на сайт, пользователь получает представление о целях создания DVM-системы, о задачах, на решение которых она ориентирована, о том, что такое DVMH-модель параллельного программирования, о существующих средствах поддержки процесса разработки параллельных программ с использованием DVMH-модели. Для начального ознакомления с ключевыми возможностями языков Fortran-DVMH и C-DVMH доступны подробно разобранные примеры программ. Примеры могут быть полезны как студентам, только начинающим изучать возможности DVM-системы так и пользователям, обладающим базовым уровнем знаний в области параллельного программирования. Для углубленного изучения DVMH-модели подготовлено отдельное руководство, ознакомившись с которым можно выполнить распараллеливание программ разного уровня сложности. Указанное руководство используется студентами при выполнении практических заданий на этапе разработки DVMH-версий последовательных программ. Проследить историю развития DVM-системы и изучить отдельные аспекты DVMH-модели можно в разделе публикаций и докладов, в котором наряду с текстами статей, опубликованными в различных журналах и материалах конференций, доступны презентации и видео с выступлений разработчиков DVM-системы на конференциях, посвященных параллельному программированию.

Получить доступ к DVM-системе можно тремя способами: установить систему на своем компьютере, получить удаленный доступ к рабочей станции или воспользоваться суперкомпьютерами, на которых система предустановлена. Персональные компьютеры и рабочие станции оптимальны для начального обучения технологиям параллельного программирования. На них нет очередей задач и ограничений по запуску программ. При этом на начальном этапе не требуется большого количества вычислительных ресурсов. Чтобы получить доступ к рабочей станции с установленной DVM-системой, необходимо заполнить форму доступа на сайте. Также на сайте доступны инструкции по использованию системы на суперкомпьютерах «Ломоносов» (НИВЦ МГУ), K100 (ИПМ им. М.В. Келдыша), МВС-10П (МСЦ РАН), МВС-100k (МСЦ РАН).

На основе вышесказанного можно сделать следующий вывод. Интернет-технологии позволяют обеспечить оперативную связь студентов с преподавателем, постоянный контроль за ходом образовательного процесса,

доступ к Интернет-ресурсам, содержащим актуальную и оперативно обновляющуюся информацию по соответствующей тематике. При обучении суперкомпьютерным технологиям для практической работы необходимо иметь доступ к вычислительным ресурсам, использование которых возможно только удаленно. Необходимо учитывать, что Интернет-ресурсы предоставляют доступ к обширному набору неструктурированной информации, достоверность источников, которой может вызывать сомнения.

В статье было рассмотрено применение сетевых технологий в процессе обучения параллельному программированию на факультете ВМиК МГУ им. М.В. Ломоносова. Студенты получают навыки удаленной работы с вычислительными ресурсами, знакомятся с правилами работы на суперкомпьютерных системах, обучаются работе с системами очередей для запуска своих программ. Обучение включает в себя знакомство с технологиями параллельного программирования MPI, OpenMP, OpenACC и DVMH. Авторами статьи был разработан сайт DVM-системы, содержащий подробное руководство по использованию системы, примеры программ, публикации и доклады. Была организована система контроля процесса обучения на базе веб-приложения для управления проектами.

## ЛИТЕРАТУРА

1. Пресс-выпуск №2836 [Электронный ресурс].  
URL: <http://wciom.ru/index.php?id=236&uid=115255> (дата обращения 14.10.2015).
2. Википедия: Отказ от ответственности [Электронный ресурс].  
URL: [http://ru.wikipedia.org/wiki/Википедия:Отказ\\_от\\_ответственности](http://ru.wikipedia.org/wiki/Википедия:Отказ_от_ответственности) (дата обращения 14.10.2015).
3. Цымблер М.Л. Big Data: несколько простых вопросов о сложном явлении // Суперкомпьютеры 1(17) весна 2014 — ООО «Издательство СКР Медиа», С. 8-11.
4. Бахтин В.А., Бородич И.Г., Катаев Н.А., Клинов М.С., Ковалева Н.В., Крюков В.А., Поддерюгина Н.В. Диалог с программистом в системе автоматизации распараллеливания САПФОР // Вестник Нижегородского университета им. Н.И. Лобачевского. — 2012 — № 5 (2). — С. 242–245.
5. OpenMP Application Program Interface Version 4.0 - July 2013 [Электронный ресурс].  
URL: <http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf> (дата обращения 14.10.2015).
6. Спецсеминар «Операционные системы и языки программирования распределенных вычислительных систем» [Электронный ресурс].  
URL: <http://sp.cs.msu.ru/specsem/ospl/> (дата обращения 14.10.2015).
7. Коновалов Н.А., Крюков В.А., Михайлов С.Н., Погребцов А.А. Fortran DVM - язык разработки мобильных параллельных программ // Ж. "Программирование", № 1, 1995, С. 49-54.



8. Бахтин В.А., Клинов М.С., Крюков В.А., Поддерюгина Н.В., Притула М.Н., Сазанов Ю.Л. Расширение DVM-модели параллельного программирования для кластеров с гетерогенными узлами. – Вестник Южно-Уральского государственного университета, серия "Математическое моделирование и программирование", №18 (277), выпуск 12 – Челябинск: Издательский центр ЮУрГУ, 2012, с. 82-92.
9. Overview – Redmine [Электронный ресурс].  
URL: <http://www.redmine.org/> (дата обращения 14.10.2015)
10. OpenACC Home | [www.openacc.org](http://www.openacc.org) [Электронный ресурс].  
URL: <http://www.openacc.org/> (дата обращения 14.10.2015).
11. Intel VTune Amplifier | Intel Developer Zone [Электронный ресурс]. URL: <https://software.intel.com/en-us/intel-vtune-amplifier-xe> (дата обращения 14.10.2015).
12. NAS Parallel Benchmarks [Электронный ресурс].  
URL: <http://www.nas.nasa.gov/publications/npb.html> (дата обращения 14.10.2015).
13. DVM-система | Система разработки параллельных программ [Электронный ресурс].  
URL: <http://dvm-system.org/> (дата обращения 14.10.2015).

# ОБ ОДНОМ ПОДХОДЕ К ОПРЕДЕЛЕНИЮ АВТОРСТВА МУЗЫКАЛЬНЫХ ПРОИЗВЕДЕНИЙ <sup>2</sup>

А.С. Борисенкова, Ю.С. Корухова

*Факультет Вычислительной математики и кибернетики  
Московского государственного университета имени М.В. Ломоносова*

В наш информационный век огромное количество данных их самых разных областей хранится в электронном виде. Работать с электронными библиотеками и коллекциями приходится практически любым специалистам, поэтому актуальными становятся инструменты эффективного поиска, хранения и анализа электронных документов. Не является исключением и музыкальная область: электронные нотные библиотеки стали хорошим помощником музыкантам, композиторам, исполнителям. Такие библиотеки содержат нотные записи разных форматов: от отсканированных партитур и рукописных нот до файлов, созданных специальными программами – нотными редакторами. Для нотных записей в электронном виде также требуются инструментальные средства.

В статье рассматривается задача автоматического анализа нотных записей музыкальных произведений с целью определения их авторства. Исходными данными является фрагмент нотной записи, для которого требуется определить, перу какого композитора он принадлежит с наибольшей вероятностью. Образцы произведений различных композиторов известны заранее. Решение такой задачи внесло бы вклад в усовершенствование нотных поисковых систем: представляется, что поиск мог бы быть быстрее при знании метаинформации (автора произведения), чем поиск по содержимому файлов. Также можно было бы применить метод для установления авторства анонимных партитур или нот, подписанных вымышленным именем, которые достаточно часто встречаются на просторах сети Интернет. Эксперт может на слух отличить, какой именно композитор написал то или иное музыкальное произведение, основываясь на личном опыте прослушивания музыки. Эксперт мог ранее слышать это произведение и знать, кто его автор, либо он может узнать некоторый индивидуальный почерк, характерный для определенного композитора, и по нему выделить произведение данного композитора из числа других. Однако для задач, требующих прослушивания тысяч произведений, использование экспертного подхода требует неприемлемых затрат времени, а в ряде случаев невозможно. Формализация правил, по которым эксперт делает выводы, представляется достаточно сложной задачей: зачастую, эксперт руководствуется интуицией и ранее приобретенным опытом. Последний факт

---

<sup>2</sup> Работа выполнена при поддержке РФФИ, проект 14-01-00679 А

наводит на мысль, что компьютерную систему идентификации композитора можно пытаться построить, используя методы машинного обучения.

### **Известные подходы к решению задачи автоматического определения авторства**

В работе [1] был предложен метод определения авторства нотной записи на основе цепей Маркова. Музыкальное произведение рассматривается как случайный процесс, который моделируют с помощью цепи Маркова с дискретным временем. В качестве пространства состояний берут ноты, встречающиеся в произведении. Похожесть произведений оценивается с помощью расстояния Кульбака – Лейблера между функциями распределения. Для обучения системы использована выборка из 100 произведений Моцарта и 212 произведений Гайдна для струнных квартетов. Для оценки точности применен метод скользящего контроля: одно произведение из обучающей выборки распознавалось, на остальных обучались. Максимальная точность, которая была достигнута – 68%, она является невысокой, тем более с учетом того, что выбор делался всего между двумя авторами.

Другой подход к решению задачи идентификации композитора, описанный в работе [2], основан на нейронных сетях. Музыкальное произведение представлено в виде двенадцатимерного вектора, каждая размерность которого соответствует одному из 12 тонов, и каждая координата суммирует количество соответствующих нот. Для классификации использовалась вероятностная нейронная сеть, состоящая из четырех слоев – входного слоя, слоя образцов, слоя суммирования и выходного слоя. Вектор, соответствующий произведению, подается на 12 входных нейронов и распространяется на слой образцов, который имеет столько же групп, сколько композиторов в обучающей выборке. В каждой группе столько нейронов, сколько произведений этого автора в обучающей выборке. Слой суммирования состоит из нейронов, вычисляющих вероятность принадлежности к соответствующему его номеру классу. Выходной слой из одного нейрона относит объект к тому классу, вероятность принадлежности к которому наибольшая.

Рассмотренная в [2] выборка состояла из произведений 7 композиторов (Бах, Бетховен, Брамс, Шопен, Гендель, Гайдн и Моцарт) по 50 произведений на каждого. Обучение было проведено на 35 произведениях каждого композитора, а оставшиеся 15 использованы для оценки точности классификации. Максимальная точность классификации (66,1%) была получена для произведений Генделя, однако из произведений Гайдна были правильно классифицированы только 24,3%, что показывает, что классификатор работает не случайным образом, но хотелось бы улучшить его результаты. При построении вектора, например, не учтена информация о длительностях нот, а она представляется существенной.

В обоих рассмотренных примерах задача определения авторства сведена к задаче классификации. Подобная задача решается в [3] для текстов. Утверждается, что у каждого автора есть свой индивидуальный стиль, и если предложить некоторый формальный способ его описания, то определение авторства текста (определение принадлежности к одному из заранее известных классов) станет возможным. В [3] предложено рассматривать плотность функций распределения последовательностей букв длины  $n$  –  $n$ -грамм:

$$\varphi(l_1, l_2, \dots, l_q) = \frac{(c_1, c_2, \dots, c_q)}{M}$$

где  $M$  – количество букв во всем тексте,  $c_n$  – количество повторений последовательности букв  $l_n$  длины  $n$ ,  $k$  – количество букв в алфавите языка, на котором написан текст,  $q = C_N^k$  – размерность полученного вектора. Таким образом каждый текст можно будет представить в виде вектора в многомерном пространстве. На полученном пространстве вводится норма

$$\langle \varphi_i, \varphi_j \rangle = \sum_{i=1}^q |\varphi_i(l_i) - \varphi_j(l_i)|$$

В качестве «авторского эталона» текстов рассматривается средневзвешенный вектор признаков:

$$\Phi(l_1, l_2, \dots, l_q) = \frac{\sum_{i=1}^t \varphi_i M_i}{\sum_{i=1}^t M_i}$$

где  $t$  – количество текстов автора.

При поиске по пространству текстов плотность функций распределения  $n$ -грамм (N-ПФР) текста из запроса сравнивают со средневзвешенными векторами признаков (средневзвешенными N-ПФР) авторов из библиотеки. Писатель, чей средневзвешенный вектор признаков окажется ближе всего к N-ПФР текста из поискового запроса, признается автором текста.

В библиотеке может не оказаться писателя, написавшего идентифицируемый текст. Это можно будет заметить по тому, что вектор признаков, соответствующий тексту из поискового запроса, окажется достаточно далеко от средневзвешенных векторов признаков авторов из библиотеки.

Также в [3] используется *расстояния разделения*. Это расстояние  $\rho$ , при котором в процессе обучения получается минимальная ошибка классификации. Расстояние дает еще более точный критерий определения авторства: текст с вектором признаков  $\varphi$  считается написанным  $i$ -м автором тогда и только тогда, когда расстояние  $\varphi$  до средневзвешенного вектора этого автора не превышает расстояние разделения  $\rho$ . При использовании такого критерия возможна ситуация, что вектор произведения находится ближе к средневзвешенному вектору одного из композиторов, но его расстояние разделения меньше, и рассматриваемое произведение относится к классу, находящемуся дальше, но имеющего большее расстояние разделения.

Построим на основе рассмотренного метода классификатор музыкальных произведений по композиторам. Напрямую применить метод не удастся, так как не определено понятие буквы, а оно используется для построения векторов. Также потребуется учесть особенности предметной области: произведение, записанное в другой тональности, является тем же произведением, но имеет другое символическое представление. Кроме этого, в нотной записи требуется учитывать знаки, показывающие особенности исполнения, повторы (репризы, вольты и т.п.). Далее необходимо провести обучение системы: определить средневзвешенные векторы композиторов, ввести меру и расстояние разделения.

### **Внутреннее представление нотных записей для классификации**

Входные данные метода – нотные записи, записанные в формате MusicXML [4]. Данный формат поддерживается многими программами – нотными редакторами наряду с их собственными форматами, поэтому получил достаточно широкое распространение. Файлы MusicXML являются частным случаем XML файлов, для которых зафиксирован набор тегов, описывающих музыкальные обозначения: высота нот, длительность, октава, тональность, темп и другие. В качестве «буквы» будем рассматривать для каждой пары соседних нот пару *<разность высот, отношение длительностей>*. Такое представление будет одинаковым для записи одного и того же произведения в разных тональностях и не меняется при равномерном изменении темпа и длительностей. Произведение может быть написано для нескольких инструментов, а значит – включать в себя несколько нотноносцев. Для некоторых инструментов (например, для фортепиано) партитура состоит из двух нотноносцев – для левой и для правой рук. Нотноносцы, играющиеся одновременно одним или несколькими инструментами, соединяются акколадой (фигурной или прямой соответственно). Предлагается разделить произведение на группы нотноносцев, объединенных акколадой, и рассчитывать последовательности нот для каждой из этих групп отдельно, как если бы они звучали последовательно друг за другом, а не все одновременно.

При построении внутреннего представления следует учесть встречающиеся в произведении репризы, обозначающие повторяющуюся часть произведения, и вольты, содержащие в себе ноты, исполняемые при повторе соответствующего номера

Чтобы посчитать количество  $n$ -грамм, для каждого из  $n$  подряд идущих списков пар в произведении выберем возможные перестановки на  $n$  идущих подряд множествах нот. Для вычисления окончательного вектора признаков (ненормированного) достаточно подсчитать для каждой  $n$ -граммы число ее появлений в произведении. Если какая-то  $n$ -грамма в произведении не встречается, принимаем ее значение равным нулю.

Рассмотрим, каково максимальное потенциальное количество  $n$ -грамм. Если принять, что в одной группе нотноносцев максимальная разность октав не

будет превосходить единицы, то может быть всего 13 различных значений высот по целым нотам или 25 различных значений высот с учетом полутонов, 9 различных значений длительностей нот или 27 различных значений длительностей с учетом точек (одинарных и двойных). Это означает, что максимальное потенциальное количество n-грамм равно:

$$(25 \cdot 2 \cdot 27 \cdot 2)^n = 2700^n$$

С ростом n размерность вектора признаков растет экспоненциально. Однако если вспомнить о смысле признаков, то обнаружим, что некоторые из них не встретятся не только в рассматриваемом произведении, но и во всей обучающей выборке. Так, например, для n=1 признак  $(-26, \frac{1}{16})$ , соответствующий повышению высоты на три октавы и шесть тонов и уменьшению длительности в 16 раз, скорее всего, не будет встречаться, тогда как признак (1,1), соответствующий уменьшению высоты ноты на 1 тон с сохранением длительности, будет встречаться в векторах произведений повсеместно.

### Обучение классификатора

Пусть имеется обучающая выборка произведений N композиторов, у каждого из которых будет  $M_i$  произведений ( $M_i > 1$ ). Для вычисления расстояния между двумя векторами признаков  $\varphi_i$ ,  $\varphi_j$ , соответствующих двум произведениям:

$$\langle \varphi_i(i), \varphi_j(i) \rangle = \sum_{i=1}^q |\varphi_i(i) - \varphi_j(i)|$$

Построим распределение отклонений N-ПФР своих произведений  $g_c^+$  и произведений других композиторов  $g_c^-$  из базы от средневзвешенного вектора (N-ПФР) рассматриваемого композитора. При вычислении  $g_c^-$  используется введенная ранее формула вычисления N-ПФР, а для  $g_c^+$  необходимо для каждого произведения  $i$  вычислять отдельно средневзвешенную N-ПФР без этого произведения и измерять расстояние от получившейся средневзвешенной N-ПФР до произведения:

$$\Phi_{g_c^+}(l_1, l_2, \dots, l_q) = \frac{\sum_{j=1, j \neq i}^t (\varphi_j \cdot M_j)}{\sum_{j=1, j \neq i}^t M_j}$$

Остановимся на задаче вычисления распределения расстояний и рассмотрим её чуть более подробно. При измерении расстояния от произведений до средневзвешенного вектора (средневзвешенной N-ПФР) получим множество вещественных значений из q элементов  $\{\rho_1, \dots, \rho_q\}$ :

$$\forall i, j \in \{1, \dots, q\}, \rho_{min} \leq \rho_j \leq \rho_{max}, \rho_j \in R$$

Требуется разбить отрезок  $[\rho_{min}, \rho_{max}]$  на интервалы так, чтобы отнести все расстояния, попавшие в один интервал, к некоторому среднему значению. В

работе [6] приводится следующая оценка количества интервалов, зависящая от числа точек  $N$ :

$$n = 4 \cdot \left(\frac{3}{4} \cdot (N - 1)^2\right)^{\frac{1}{5}}$$

Будем пользоваться этой оценкой для вычисления оптимального количества интервалов при разбиении отрезка. В качестве расстояния разделения  $\rho$  примем такое значение  $\rho$ , при котором ошибка идентификации композитора минимальна:

$$\rho = \arg \min_{\rho} (1 - G_c^+(\rho) + G_c^-(\rho))$$

где  $G_c^+(\rho)$  и  $G_c^-(\rho)$  – соответствующие  $g_c^+$  и  $g_c^-$  интегральные функции распределения.

Таким образом, вокруг средневзвешенного вектора произведений рассматриваемого композитора построена многомерная сфера с радиусом  $\rho$ , все произведения внутри которой будут относиться к текущему композитору, а произведения снаружи этой сферы будут считаться чужими. Если вектор произведения из поискового запроса будет располагаться далее, чем на  $\rho$  от средневзвешенных векторов всех композиторов (вообще говоря,  $\rho$  свое для каждого композитора), то будет принято решение, что композитора, написавшего искомое произведение, в базе нет. Однако в случае, когда вектор произведения–запроса находится от всех средневзвешенных векторов на расстоянии, большем всех расстояний разделения, предлагается относить данное произведение к тому композитору, чей средневзвешенный вектор ближе. Аналогичный подход применялся и в [3] для текстов.

Для вектора–запроса оцениваются сначала расстояния до средневзвешенных векторов известных композиторов, попадающие в сферы с заданным расстоянием разделения, и в формулируемом ответе на запрос соответствующие фамилии композиторов упорядочиваются по возрастанию расстояний. После них в список добавляются (в порядке возрастания расстояний) остальные композиторы.

### **Реализация и оценка работы метода**

Предложенный метод был реализован в программной системе на языке Ruby, информация о произведениях, на которых проведено обучение, хранится в базе данных СУБД PostgreSQL. Для экспериментов были составлены 10 обучающих выборок векторов признаков для  $N = \{1,2,3,4,5\}$  из произведений 19 композиторов (в одной выборке использовано максимум 16 композиторов). Для большинства выборок максимальное значение F-меры (порядка 0,8) было получено при  $N=5$ .

Для оценки работы алгоритма, с использованием библиотек Nimbus и Parallel был также реализован классификатор методом случайного леса [5]. Для построения решающих деревьев использован алгоритм ID3 с индексом Джини в качестве меры неоднородности. В результате экспериментов было получено,

что с ростом  $N$  F-мера предложенного в работе метода растет, тогда как при использовании метода случайного леса она уменьшается.

Для выборки, содержащей произведения максимального количества композиторов (16), были получены результаты, представленные в таблице 1.

	Точность	Полнота	F-мера
Моцарт	0.95	1.0	0.97
Бах	0.9	1.0	0.95
Шуман	0.9	1.0	0.95
Гендель	0.85	1.0	0.92
Верди	1.0	0.83	0.91
Шопен	0.9	0.9	0.9
Гайдн	0.7	1.0	0.82
Брамс	0.65	1.0	0.79
Штраус	0.65	1.0	0.79
Мендельсон	0.65	1.0	0.79
Дебюсси	1.0	0.61	0.75
Россини	0.65	0.87	0.74
Бетховен	0.5	1.0	0.67
Вивальди	0.8	0.53	0.64
As I Lay Dying	1.0	0.36	0.53
Чайковский	0.35	1.0	0.52

Таблица 1. Оценка метода при классификации по 16 композиторам

Метод решающих деревьев показал F-меру 0.23 для  $N=1$ . Для  $N>1$  обучение занимало значительное количество времени, а точность классификации при этом уменьшалась.

Однако метод случайного леса представляет особый интерес при оценке важности признаков вектора, по которым выполняется классификация, так как построение решающего дерева может завершиться (и часто завершается) до того, как были использованы все признаки. Выделив признаки, использованные при построении дерева, мы рассматриваем вычисление расстояния только для координат, соответствующих этим признакам.

### Заключение

В работе предложен подход к определению авторства нотных записей музыкальных произведений. Метод реализован в программной системе, которая решает задачу классификации нотных записей в формате MusicXML по композиторам. Система включает в себя модуль перевода нотной записи в специальное внутреннее представление, ведение базы данных по композиторам и их произведениям, обучение и оценивание классификаторов, построение



графиков, представляющих произведение из запроса в многомерном признаковом пространстве.

Предложенный метод показал оценку F – меры порядка 0.8 для 16 классов. Список композиторов может быть расширен при проведении дополнительного обучения системы. Предложенный в работе подход может также быть применен и для решения других задач классификации музыкальных произведений: по жанрам, стилям, музыкальным направлениям.

## ЛИТЕРАТУРА

1. Yi-Wien Liu Modelling Music as Markov Chains: Composer Identification – URL: [http://esf.ccarh.org/254/254\\_LiteraturePack1/ComposerID\\_Liu.pdf](http://esf.ccarh.org/254/254_LiteraturePack1/ComposerID_Liu.pdf)
2. Maximos A. Kaliakatsos-Papakostas, Michael G. Epitropakis, Michael N. Vrahatis Musical Composer Identification through Probabilistic and Feedforward Neural Networks // Lecture Notes in Computer Science, Vol.6025, p.411-420, 2010
3. Борисов Л.А., Орлов Ю.Н., Осминин К.П. Идентификация автора текста по распределению частот буквосочетаний // Препринты ИПМ им. М.В.Келдыша. 2013. № 27. 26 с. URL: <http://library.keldysh.ru/preprint.asp?id=2013-27>
4. Формат MusicXML URL: <http://www.musicxml.com/>
5. Yali A., German D. Shape Quantization and Recognition with Randomized Trees // Neural Computation, Vol. 9, p.1545-1588, 1997
6. Орлов Ю.Н. Оптимальное разбиение гистограммы для оценивания выборочной плотности функции распределения нестационарного временного ряда // Препринты ИПМ им. М.В.Келдыша. 2013. № 14. 26 с. URL: <http://library.keldysh.ru/preprint.asp?id=2013-14>

# **ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ПО КРИСТАЛЛОХИМИИ И МИНЕРАЛОГИИ WWW-MINCRYST: 18 ЛЕТ НАУЧНОМУ ИНТЕРНЕТ-ИНСТРУМЕНТАРИЮ**

**Д.А. Варламов<sup>1</sup>, Т.Н. Докина<sup>1</sup>, Н.А. Дрожжина<sup>1</sup>, О.Л. Самохвалова<sup>1</sup>**

<sup>1</sup> *Институт экспериментальной минералогии РАН, г. Черноголовка, РФ*

Информационно-вычислительная система (ИВС) WWW-MINCRYST (<http://database.iem.ac.ru/mincryst>, <http://mincryst.iem.ac.ru>), предназначенная для работы с максимально возможным количеством пользователей и поэтому изначально базирующаяся на использование интернет-технологий, в качестве базы данных, размещенной в среде Интернет, возникла достаточно давно и стала одним из пионерских интерактивных научных Интернет-ресурсов как в России, так и в мире (в области наук о Земле). Первый полностью работоспособный вариант WWW интерфейса был создан в рамках гранта РФФИ № 96-07-89162 и представлен пользователям в декабре 1997 года [1], таким образом система успешно функционирует и активно востребована пользователями уже почти 18 лет.

Ниже описывается созданная в 1997 году и активно развиваемая вплоть до настоящего времени ориентированная на Интернет-пользователя информационно-вычислительная система (ИВС) WWW-MINCRYST со свободным доступом. Цель создания ИВС WWW-MINCRYST с ее обширным инструментарием – обеспечить максимально широкий круг пользователей во всех областях науки, оперирующих с кристаллическим веществом (геология, геохимия, кристаллография, физика твердого тела) достоверной и актуальной информацией в области кристаллохимии и кристаллографии минералов, их синтетических аналогов и элементов, а также дать дополнительный толчок в развитии методов обработки и визуализации кристаллоструктурных данных.

Первоначально (в конце 90-х) ресурс был призван обеспечить интерактивный доступ Интернет-пользователям к накопивавшимся с 1985 года в рамках локальной ИВС МИНКРИСТ данным по кристаллическим структурам (прежде всего литературным, а также авторским аналитическим — всего около 3500 объектов на момент старта проекта). В ходе проекта в состав ИВС вводились программные разработки по обработке кристаллоструктурных и кристаллохимических данных, их анализу и мультимедийной визуализации. Главным инициатором и основным идеологом работ был заведующий группой рентгеноспектрального анализа Института экспериментальной минералогии РАН выдающийся кристаллограф А.В. Чичагов (увы, безвременно ушедший от нас в 2010 году).

За 18 лет развития ИВС WWW-MINCRYST нарастил свой информационный фонд почти в 2,5 раза (почти на 6000 объектов) и по праву вошел в первые ряды рентгеноструктурных и кристаллохимических баз

данных, связанных с изучением минерального вещества, как по оценкам пользователей, так и по мнению составителей отраслевых каталогов информационных ссылок.

ИВС WWW-MINCRYST зарегистрирована в государственном регистре баз данных (НТЦ "Информрегистр") под номером 0229805169 (регистрационное свидетельство № 4873 от 11.02.99).

Основным компонентом ИВС является собственно база данных или Информационный фонд (на июнь 2015 года — около 9300 информационных объектов для почти 4000 уникальных фаз, включая 3700 природных минералов и около 300 синтетических фаз, не имеющих пока натуральных аналогов). Информационный объект состоит из 4 базовых таблиц с 60 фиксированными и до 130 дополнительными полями (число последних варьирует в зависимости от сложности структуры) и 4 бинарных файлов. Поддерживается иерархическая система индексов и ряд вспомогательных таблиц. Общий объем базы данных составляет около 500 Мбайт.

Информационный фонд в ИВС WWW-MINCRYST содержит информацию о большинстве минеральных видов (около 3700 из почти 5000 официально признанных на май 2015 года Международной Минералогической Ассоциацией – <http://www.ima-mineralogy.org/Minlist.htm>), кристаллические структуры которых расшифрованы к настоящему времени. Помимо природных объектов, в базе данных представлены синтетические минералы – их структурные аналоги, отличающиеся по составу (например, с заменой одного из катионов), и неорганические соединения (силикаты, фосфаты, бораты и др.), близкие по свойствам к природным веществам. Информационный фонд содержит данные структурных работ из более 130 иностранных и отечественных журналов за период от 1930-х годов по настоящее время. Ежегодный прирост новых или заново переопределенных/уточненных кристаллических структур минералов и их аналогов достаточно значителен, чтобы требовалась постоянная актуализация информационного фонда (в среднем, до 350-450 структур в год). В последние годы основной акцент был сделан на новые минералы, а также на кристаллические структуры, описанные в советских и российских журналах (которые в последние десятилетия были менее доступны web-пользователю, нежели зарубежные).

Интерфейс ИВС WWW-МИНКРИСТ представляет собой многоуровневую информационную двуязычную (русский и английский, включая содержание и языки интерфейса) систему. В нее входят следующие компоненты: (а) комплексные поисковые интерфейсы, использующие как критерии поиска (в различных комбинациях) названия минералов, химический состав, кристаллоструктурные параметры, литературные ссылки и вспомогательную информацию, причем возможны комбинации поисковых признаков; (б) модули мультимедийного представления информации (интерактивные полиэдрические и шаровые структуры, линейчатые и непрерывные спектры — через Java-апплеты WWW-CrystPic и WWW-MixiPol;

(в) классификационные схемы; (г) модули организации взаимосвязей с внешними информационными ресурсами через систему динамически формируемых ссылок; (д) WWW-ориентированный инструментарий разработчика (включая систему импорта, проверки и редакции данных, а также архивации и резервного копирования).

Технологически ИВС WWW-MINCRYST как web-сервис реализована на "классической" связке Linux-Apache-Mysql-PHP (LAMP) с использованием JavaScript и (для интерактивных апплетов) Java (на базе Sun/Oracle JDK). Размещена ИВС на сервере баз данных Институт экспериментальной минералогии РАН (<http://database.iem.ac.ru>), доступ нелимитирован. На старте проекта ИВС WWW-MINCRYST стала в 1999 году (в реализации на ОС Digital Unix) одной из первых интернет-ориентированных баз данных в России, использовавшим данную технологию, причем за 15 лет она нисколько не потеряла свою технологичность на стороне сервера. В 2015 проведена миграция базы данных с проприетарной СУБД MySQL на свободно распространяемую MariaDB (<https://mariadb.org/>) для сохранения свободного статуса ИВС. В настоящее время ИВС размещена на сервере со следующими параметрами – CPU Intel Core i5-2400@3.10GHz x 4 ядра, RAM 24 Gb, disk array SSD 128 Gb + 2x2 Tb WD RE 2000, GPU Nvidia Kepler GTX 770, программно - Apache HTTPD server 2.4.6, PHP script language 5.4.16, MariaDB database 5.5.40.

В теоретической основе ИВС WWW-MINCRYST лежит ряд принципиальных "идеологических" моментов, реализация которых и обеспечила ему имеющееся признание кристаллографической и минералогической научной общественности. Синтез информации о кристаллической фазе, рассматриваемой как "монокристалл" и/или как "поликристалл" с заменой экспериментальных поликристалл-стандартов расчетными – базовая идея WWW-MINCRYST, которая и делает его универсальной, комплексной и оригинальной разработкой, поскольку позволяет существенно расширить информационную базу сравнительно с чисто аналитической. На базе введенной информации программный пакет эксперта-разработчика позволяет автоматически сформировать вторую, производную от первой (аналитической), базу расчетных поликристалл-стандартов, проводя синтез двух типов информации о кристаллической фазе. Связка "Кристаллическая структура фазы и ее расчетная поликристалл-рентгенограмма" является информационной основой ИВС WWW-MINCRYST и служит важнейшим инструментарием в руках пользователя (в особенности рентгеновского кристаллографа). Стержневая идея WWW-MINCRYST о синтезе двух типов кристаллоструктурной информации о кристаллической фазе реализует принципиально новый подход к формированию всей кристаллоструктурной информации о веществе и организации доступа к ней.

Основным источником информации для ИВС служат оригинальные журнальные статьи, опубликованные в открытой печати. Извлеченная кристаллоструктурная информация помещается (в соответствии со

специальным форматом записи) в ASCII-файл с последующей программной экспертизой (проводимой на локальных ПК) по результатам расчета межатомных расстояний и других кристаллоструктурных характеристик и, в случае положительного решения, импортируется в ИВС специальными средствами группового или одиночного импорта. Возможен импорт кристаллоструктурной информации через общепринятый в кристаллографии специализированный файл (Crystallographic Information File, CIF), содержащий кристаллоструктурные характеристики расшифрованной кристаллической фазы (формат – <http://www.sdsc.edu/pb/cif/cif.html>).

Информационная запись для индивидуального кристаллического вещества содержит информацию о названии (в соответствии с классификацией International Mineralogy Association или рекомендациями по наименованию неорганических веществ IUPAC), химическом составе, симметрии, параметрах элементарной ячейки, координатах атомных позиций с изотропными температурными факторами и заселенностями, информацию о межплоскостных расстояниях, HKL-индексах и интенсивностях сильнейших рефлексов рентгенодифракционной картины поликристалл-фазы, а также ссылки на соответствующие публикации по расшифровке или уточнению кристаллической структуры. Запись может быть специфицирована по полезным свойствам, особенностям химического состава и структуры, P-T условиям синтеза, принадлежности к условным минеральным группам. Каждая запись содержит "монокристалльные" и "поликристалльные" характеристики кристаллической фазы. Минералы классифицированы в соответствии с таксонами структурно-химической систематики минералов А.А. Годовикова, кристаллохимической классификации М. Чириотти, в настоящее время вводится классификация по структурным типам минералов по Г.Б. Бокию. Для 2400 фаз сделаны экспресс-оценки потенциальной энергии кристаллической решетки. В настоящее время создается система расчета содержаний элементов в соответствии с идеальной и реальной формулами минерала-образца.

В ИВС WWW-MINCRYST реализован импорт одобренной экспертами информации через web-интерфейс как в виде единичных записей, так и пакетов записей (до нескольких сот) с входным контролем данных, что позволяет проводить постоянную актуализацию информационного фонда. Также реализована возможность редактирования записей в on-line режиме, их удаления, замены служебных файлов записей, архивации и восстановления базы данных через web-интерфейс.

Кратко об основных возможностях ИВС WWW-MINCRYST.

В ИВС реализована достаточно эффективная система поиска фаз по отдельным наборам или комплексу параметров: название минерала (полное или частичное, по спецификации, в составе общеупотребительных групп типа цеолитов, алфавитные списки), химический или элементный состав в различных комбинациях (присутствие/отсутствие элементов и их комбинации), кристаллоструктурные характеристики (симметрия, пространственные группы,

параметры элементарной ячейки, межплоскостные расстояния  $d(hkl)$ ), что в сочетании с химическим (элементным) составом дает возможность прямого интерактивного качественного рентгенофазового анализа. В систему поиска добавлен поиск по классификационным параметрам нескольких кристаллохимических и структурных классификаций (Годовиков, Чириотти, Бокий). Система поиска по этим параметрам обеспечивает поиск (и группировку) минералов по указанным низшим и промежуточным таксонам классификаций. По ряду параметров поиска ИВС WWW-MINCRYST до сих пор не имеет аналогов среди минералого-кристаллографических баз данных.

По результатам поиска пользователь получает ссылку на список фаз, отвечающих поисковым критериям с указанием ID записи, названия фазы, формулы и пространственной группы с возможностью последующего перехода по гиперссылке. В случае единичного успешного ответа пользователь сразу автоматически переводится на найденный объект.

В «карте» найденного объекта представлены основные данные (как хранимые в БД, так и динамически рассчитываемые при формировании «карты») – название, формула, параметры ячейки, различные кристаллографические данные (кол-во рефлексов, рентгеновские плотности, коэффициенты поглощения); CPDS карта – 20 максимальных рефлексов, их  $hkl$  позиции и интенсивности; базовые атомные позиции и их заселенности; полная информационно-расчетная карта (рентгеновские плотности, все рефлексы и т.п.); рассчитанные энергии решетки; принадлежность к различным классификационным схемам; источники данных (ссылки на публикации, поправки и т.п.); автоматически формируемые ссылки на внешние ресурсы.

Для всех записей через апплет WWW-Crystpic (Java-3D с использованием библиотек OpenGL) доступны динамически создаваемые интерактивные изображения моделей кристаллических структур в шарах-сферах и в полиэдрических проекциях (до 138 позиций и до 1500(!) атомов на структуру) в соответствии с основными канонами минералогии и кристаллографии. Программа позволяет делать всевозможные манипуляции с моделью структуры, включая масштабирование, непрерывное и/или автоматическое дискретное вращение вокруг "экранных" осей X,Y,Z; ориентацию по кристаллографическим осям,  $hkl$ -фрагментацию структуры (на  $hkl$ -ориентированные фрагменты толщиной  $d(hkl)$ ), наращивание элементарных ячеек вдоль любых выбранных направлений для формирования "сверхструктур" и мотивов, а также прямой "ручной" и автоматизированный для малых полиэдров (тетраэдров и октаэдров) расчет любых межатомных расстояний и углов (плоских и телесных) в структуре. Программа изображает любые полиэдры, включая "дефектные" с необычно малыми ("плохими") межатомными расстояниями. Более детальное описание всех возможностей апплета приведено здесь – <http://mincryst.iem.ac.ru/rus/crystpic.php>. В настоящее время существуют проблемы с работой с внешними библиотеками OpenGL (на которых реализован апплет) и параноидальными настройками безопасности

новых версий Java, которые решаются вмешательством пользователя. Однако, из-за этого сейчас необходим перенос апплета на новые программные платформы типа встроенных средств HTML5. Примеры моделей кристаллических структур цеолитов (до 600 атомов) приведены на рис.1

Модуль в виде апплета WWW-Mixipol предназначен для графического представления полных расчетных спектральных профилей поликристалл-рентгенограмм с возможностями манипулирования спектрами для разных источников излучения и разных типов спектральных шкал. Также модуль способен формировать рентгенограммы смесей фаз (до 6 фаз одновременно) при возможности варьирования относительными содержаниями компонентов смеси. Пример расчетных спектров смеси шести фаз приведены на рис.2

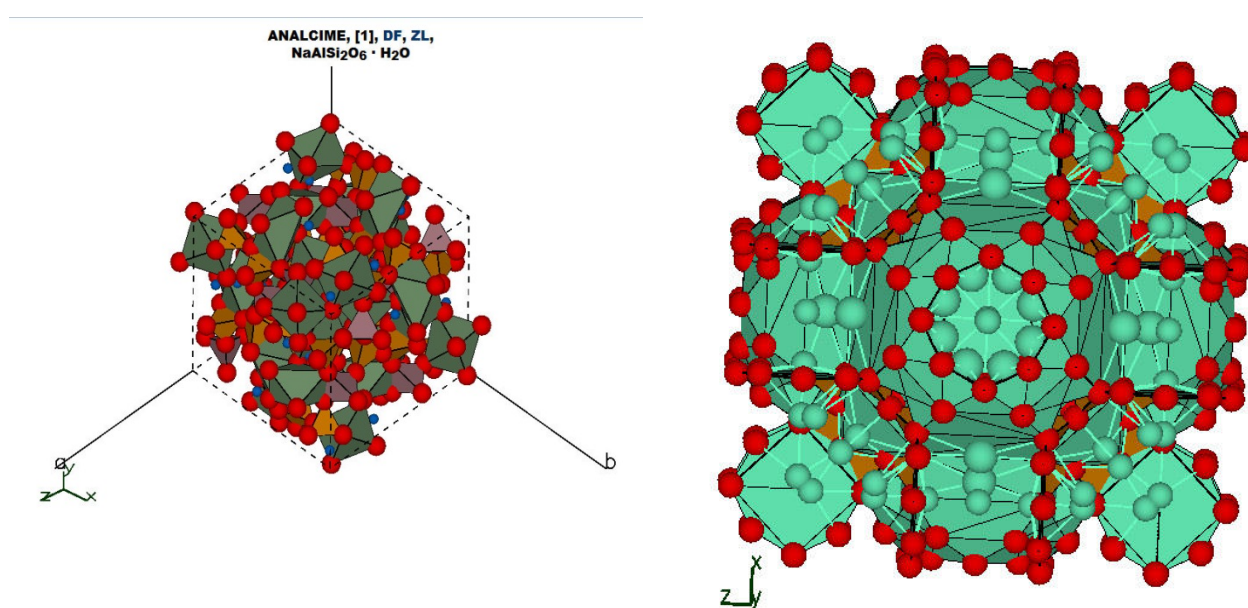


Рис. 1. Примеры интерактивных моделей кристаллической структуры (цеолиты разных типов, смешанное представление в шарах и полиэдрах, размеры атомов уменьшены вдвое)

Как для структур, так и для спектров минералов предусмотрены и упрощенные варианты представления в виде так же интерактивных традиционных шаровых структур и линейчатых спектров.

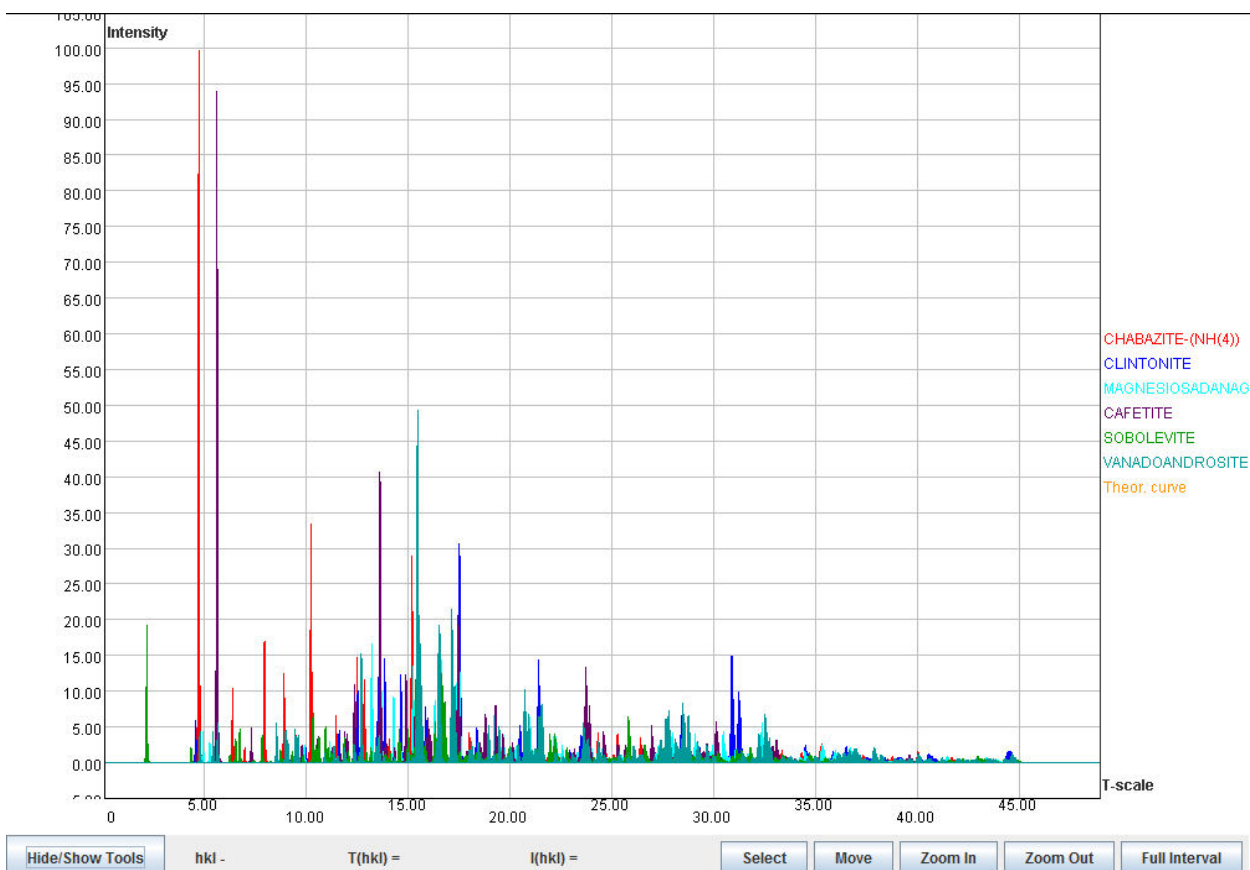


Рис.2. Пример комбинации расчетных спектральных профилей поликристалл-рентгенограмм (смесь 6 фаз)

Разработанный и интегрированный (пока частично) в WWW-MINCRYST расчетный модуль WWW-Хraypol в связке с модулями WWW-Crystpic и WWW-Mixipol позволяет квалифицированному пользователю извлекать и модифицировать "на лету" структуры из информационного фонда (не затрагивая его непосредственно), а затем рассчитывать новые варианты и получать графические изображения моделей кристаллических структур и квазиреальных полных профилей расчетных поликристалл-рентгенограмм, тем самым решая задачи по модификации реальных структур.

Одними из первых среди научных интернет-ориентированных баз данных для ИВС WWW-MINCRYST была разработана система динамически формируемых перекрестных веб-ссылок для связи объектов с записями для конкретных минералов в ведущих минералогических базах данных в Интернете. Система генерации динамических гиперссылок на внешние информационные ресурсы (на минералогические базы данных и поисковые системы) позволяет "прозрачно" для пользователя подключать большие внешние массивы данных, используя метод "генеральных" запросов. При этом пользователь сразу получает доступ к информации по интересующему его объекту, минуя стадии поиска или просмотра всей внешней базы. Кроме того, данный механизм реализует обратную связь, позволяя таким же образом



ссылаться этим базам уже на наши информационные объекты, что резко повышает востребованность ИВС внешними пользователями.

Наличие в ИВС WWW-MINCRYST данных по 9300 кристаллическим структурам, встроенного расчетного комплекса и развитых средств визуализации позволили помимо возможностей, ориентированных на поиск и предоставление информации, использовать ИВС в разработке нетрадиционных научных подходов к интерпретации и представлению некоторых кристаллических структур. Например, WWW Xraypol позволил выявить в традиционных структурах возможность гибкого использования полиэдров и почти автоматически формировать различные варианты структурных моделей минералов. Как выяснилось благодаря гибкому использованию полиэдров, позволяющему формировать различные варианты структурных моделей минералов, для части минералов можно не строго привязываться к традиционному катионно-анионному изображению, а формировать структуры на основе любых атомов, входящих в ее состав. Используется принцип: в ряде случаев кристаллическое пространство можно организовать в смешанном шаровом и полиэдрическом изображении на основе любых (а не общепринятых) атомов в структуре. Метод особенно эффективен для сложных "неправильных" бескислородных структур (фосфиды, сложные сульфиды и др.).

Востребованность WWW-MINCRYST хорошо подтверждается статистикой обращений (за 2014 год – 7,3 млн. успешных единичных запросов, более 77 Гб скачанной информации, около 48000 уникальных сайтов-клиентов), а также большим количеством отзывов, описаний и внешних ссылок на WWW-MINCRYST (см. раздел "Ссылки" на сайте).

О перспективах развития WWW-MINCRYST следует сказать отдельно. В связи с очень быстрым развитием web-технологий и (соответственно) браузеров, возникла необходимость кардинальной переработки нынешнего клиентского интерфейса, уже не отвечающего как новым технологиям (и зачастую не полностью совместимого с новым ПО), так и требованиям пользователей. Она будет включать изменение способа представления записи путем сведения разобренных сейчас полей (основные данные, CPDS карта, атомные позиции, структура, спектры, ссылки и др.) в единое информационное пространство, формируемое на базе динамического HTML (с использованием технологий JSON, DHTML, HTML5, оверлейных структур). Предусматривается как переделка и рекомпиляция Java апплетов (в связи с устареванием и неполной совместимостью кода с последними реализациями Java VM), так и их замена на модули, использующие мультимедийные технологии стандартов HTML5 или Adobe Flash (с использованием оригинального авторского расчетного кода). Также будут изменены в сторону повышения дружелюбности и простоты) дизайн и элементы управления предоставляемой пользователю информации. Предусмотрено пополнение базы данных дополнительной сугубо минералогической информацией (фото минералов, рисунки

кристаллографических форм, дополнительные минералогические данные и др.). Несмотря на то, что по ряду параметров поиска WWW-MINCRYST до сих пор не имеет аналогов, будет расширен круг потенциальных поисковых запросов (прежде всего в области поисков по составу и кристаллографическим данным).

Разработанная и развиваемая на протяжении 18 лет ИВС WWW-MINCRYST является общедоступным, дружелюбным пользователю интерфейсом к большому объему кристаллоструктурной и кристаллохимической информации с развитыми средствами поиска, представления и обработки и может служить мощным инструментарием для всех исследователей в минералогии, кристаллографии, физике твердого тела, материаловедении и прочих смежных областях науки.

Проведение работ по ИВС WWW-MINCRYST в течение 1997-2014 годов было поддержано 6-ю грантами РФФИ (в том числе в настоящее время – грантом РФФИ № 15-07-08399-а, рук. Варламов Д.А.).

## ЛИТЕРАТУРА

1. А.В. Чичагов, Д.А. Варламов, Е.В. Ершов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова Кристаллографическая и кристаллохимическая база данных для минералов и их структурных аналогов (WWW-MINCRYST) // Записки РМО, 2007, т.136, № 3, с.135-141
2. Д.А. Варламов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова WWW-MINCRYST: Интернет-ориентированная информационно-вычислительная система по кристаллографии и кристаллохимии минералов // Вестник ЮУрГУ, Сер. «Вычислительная математика и информатика», 2013, т.2, вып.1, с.26-32
3. Д.А. Варламов, Т.Н. Докина, Н.А. Дрожжина, О.Л. Самохвалова WWW-MINCRYST: информационно-вычислительная система по кристаллохимии и минералогии — техническая реализация, возможности, перспективы // Труды XVI Международной суперкомпьютерной конференции «Научный сервис в сети Интернет: многообразие суперкомпьютерных миров», 22-27 сентября 2014 года, Новороссийск – М.: Изд-во МГУ, 2014. (497 с.) с.402-405

# СИСТЕМА «ИСТИНА» ДЛЯ ПОДГОТОВКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ АНАЛИЗА НАУКОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ

В.А. Васенин, С.А. Афонин, А.С. Козицын, Д.Д. Голомазов

*НИИ механики МГУ имени М.В. Ломоносова*

Проблема создания и развития информационных систем, поддерживающих «большие данные» (Big Data – systems) на настоящее время является одной из острейших. Вопросы накопления, систематизации, хранения и обработки больших объемов данных были востребованы на практике всегда. В основном на их решение были направлены простейшие каталоги, затем так называемые базы данных, которые появились еще в 1950-х годах, а также системы управления (СУБД) ими. Базы данных различной архитектуры, модели, на которых они строились (создавались) прошли большой путь - от иерархических и сетевых (конец 1950-х - 1960-е годы) до реляционных (1970-е годы). Базы данных и системы управления на основе реляционной модели их представления активно используются на практике и в настоящее время. Наиболее известные из них, и активно используемые в разных предметных областях – ORACLE DB (Data Base), Microsoft SQL Server, PostgreSQL и ряд других. Однако все такие СУБД и базы данных на их основе, как правило, имеют дело с логически связанными данными, концептуальная структура которых заранее строго определена и однородна.

Однако жизнь быстро меняется и вместе с новыми технологиями пакетных коммуникаций, с развитием сети Интернет (с 1970-х г.) и ее информационным наполнением (1990-е г.) на основе веб-технологий для конечных пользователей (в первую очередь, в Интернет – паутине) стали доступны очень большие, даже по нынешним меркам, объемы данных разной степени структурированности. К их числу относятся данные – от неструктурированных (аудио-видео) и слабоструктурированных (текстовых) до строго структурированных, которые, однако, хранятся в базах (в т.ч. – перечисленных выше - реляционных) с разной структурой. Эти данные представляют очень большой практический интерес (физика высоких энергий, геномная инженерия, аэродинамика, астрономия, анализ текстов в интересах обеспечения безопасности и других). Однако работать с ними в рамках парадигмы традиционных реляционных баз стало уже невозможно. Эти причины привели к необходимости поиска новой парадигмы, отвечающей вызовам Big Data.

Большие Данные характеризуются целым рядом свойств, в том числе следующими, которые требуют их учета и определяют дополнительные требования, которые должны учитывать разработчики новых информационных систем.

- Данные хранятся и могут поступать из разных источников (от персоналий, источников в Интернет, из различных, уже сложившихся реляционных баз и т.д.). Нужно иметь возможность (механизмы) их эффективно собирать и систематизировать для последующей первичной обработки и хранения.
- Данные, как правило, имеют разную структуру. В этой связи нужно иметь механизмы их конвертации в форму, удобную для первичной (предварительной) обработки с целью последующей верификации и организации хранения.
- Кроме того, что данные имеют разную структуру, их очень много, что приводит к сложности поиска тех из них, которые необходимы по запросу, а также к трудностям обработки. Необходимо иметь механизмы эффективного поиска и анализа данных, а значит – предварительной систематизации (кластеризации) и поиска эффективных механизмов обработки запросов.

Отмеченные выше механизмы имеют разную природу. Это и новые модели, реализующие их средства автоматизации процессов сопровождения данных, и в том числе – некоторые меры и действия административного характера.

Одним из направлений, где востребованы анализ и обработка больших объемов гетерогенных (разнородных) данных, является управление научно-технической и образовательной деятельностью. В последние 2-3 года потребность в создании информационно-аналитической системы подготовки принятия решений на основе анализа данных научно-инновационного и образовательного содержания настоятельно обозначается на всех уровнях национальной инфраструктуры управления образованием и наукой. Систему подобного назначения (далее для краткости изложения – Систему) хотели бы иметь как отдельные организации (вузы, НИИ, научно-производственные центры и т.п.), особенно крупные, претендующие на высокий национальный и международный рейтинги, так и субъекты управления в регионах, Министерство образования и науки России. Эта потребность выражается в различных грантах и проектах, нормативно-законодательных документах, выступлениях и комментариях в СМИ представителей руководящих структур разных уровней государственной власти. Основная посылка (мотив) такой потребности и интереса: создать Систему, инструментальные средства которой позволяли бы **оперативно** анализировать (по различным срезам и, соответственно, запросам) **надежно верифицируемые** данные как **персональные** и **обобщенные по отдельным научным коллективам и организациям** (вузам, НИИ и т.п.), так и данные, обобщенные **по регионам и по России в целом, с целью получения с высокой степенью объективности оценки показателей результативности и/или тенденции (тренды) научно-педагогической и инновационно-внедренческой деятельности**. В условиях социально-ориентированной экономики, в которой преобладает государственный научно-образовательный сектор, такой инструментарий

создает дополнительные механизмы стимулирующие её переход инновационный путь развития.

Полученные с помощью Системы **оценки и результаты анализа трендов** научно-технической, образовательной и инновационной деятельности могут быть положены **в основу**: различного рода **отчетной документации; принятия управленческих** (финансово-экономических, кадровых, структурных и т.п.) **решений; определения рейтинговых показателей** субъектов подобной деятельности на всех уровнях. С учетом перечисленных требований критериями готовности целевой Системы к решению поставленных перед ней задач должны быть следующие:

1. **максимально точные** («очищенные» и верифицируемые) **в должном объеме** (по отношению к численности подлежащей анализу группы - организация, регион, страна в целом) данные о научно-педагогической и инновационно-внедренческой деятельности отдельных, участвующих в ней персон;
2. эти **данные**, в первую очередь, **должны включать**, как **сведения библиографического характера**, так и **сведения** (лучше на русском и английском языках) **аннотационного характера**, **включая ссылки** на другие источники;
3. для реализации требований к системе в полном объеме **должны быть по возможности полно (в рамках авторских прав) представлены результаты деятельности персоналий в полнотекстовом виде.**

Критерий 2 позволяет проводить по базе данных Системы как аналитические исследования (выполнять запросы) по различным срезам, так и получать индексы цитирования. Эта информация является исходной для принятия управленческих решений.

Критерий 3:

- позволяет повысить рейтинг организации, т.к. объем полнотекстовых данных входит в такие показатели;
- обеспечивает рекламу и пропаганду деятельности организации вовне;
- повышает индекс цитирования как персоналий, так и организации, которую они представляют.

Решение всех перечисленных выше задач требует ресурсоемких научных исследований, как фундаментального характера, так и прикладных; практических работ по развитию и сопровождению системы; мер организационно-административного характера, способствующих вовлечению в активную работу Системы максимально возможного (как можно большего) числа персоналий – ученых и преподавателей.

Представленные далее материалы преследуют следующие цели:

- во-первых, они излагают позицию авторского коллектива по вопросам создания и развития Систем подобного назначения в России как в плане выполнения необходимых для этого научно-поисковых исследований, так и прикладных работ;

- во-вторых – иллюстрируют востребованность такой Системы и ее функциональные возможности, состояние и перспективы эволюции на примере конкретного программного комплекса «Наука МГУ» - «ИСТИНА».

Одно из направлений научно-практической деятельности Московского университета, которое тесно связано с проблемой больших наукометрических данных и с вопросами эффективного управления исследованиями на их основе, является создание и развитие Информационно-аналитической системы «Наука МГУ», которая построена на основе моделей, механизмов и инструментальных средств Интеллектуальной Системы Тематического Исследования Научно-технической информации – «ИСТИНА» (далее – Система). Не будем вдаваться в подробности ее реализации, с которыми можно ознакомиться по монографии [1]. Отметим лишь общие архитектурные особенности, принципы, модели и алгоритмы, положенные в основу разработки Системы и программно реализованные на настоящее время.

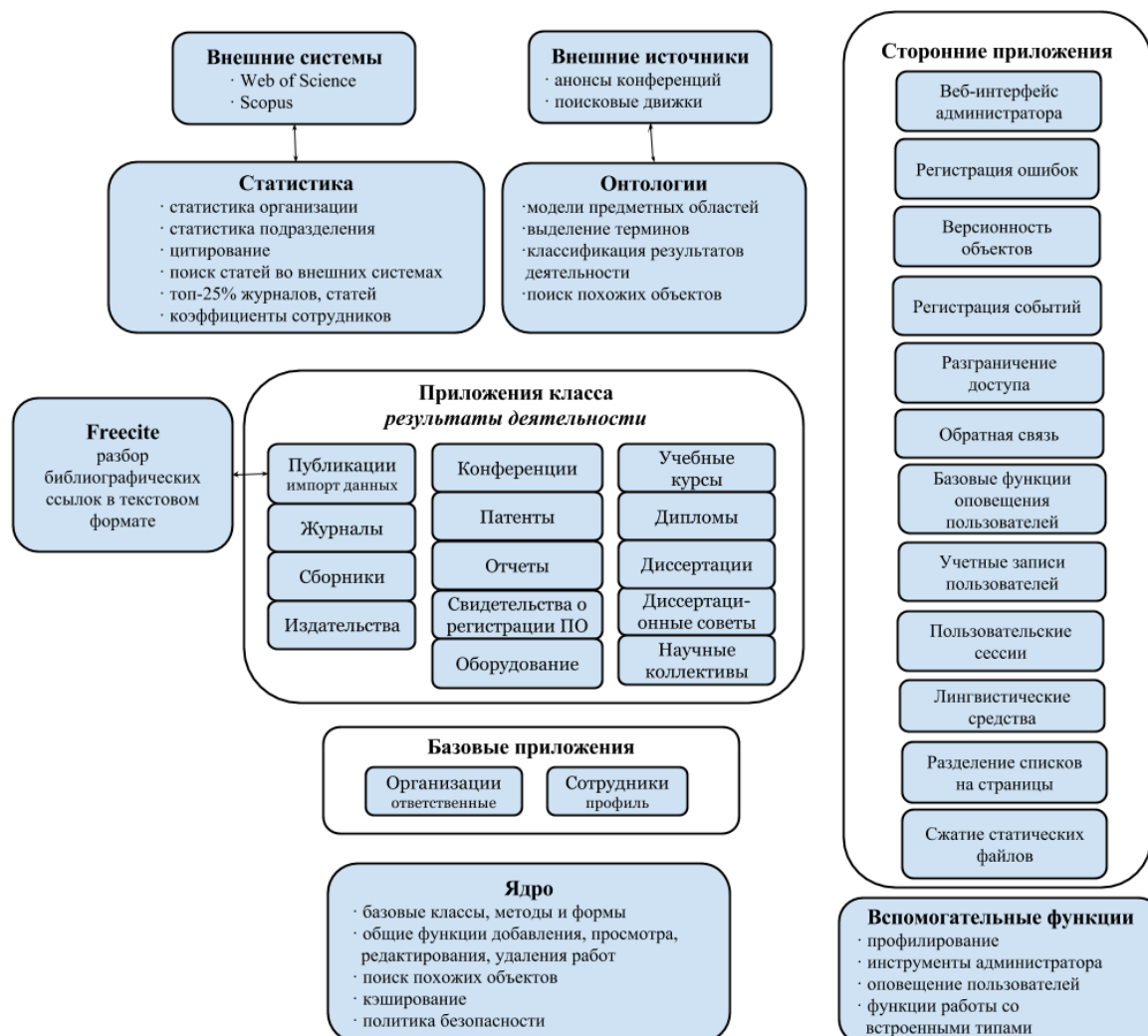


Рис. 1. Архитектура сервера приложений системы ИСТИНА.

Взаимодействие пользователей с Системой осуществляется через веб-интерфейс. Архитектура Системы является типовой для современных веб-приложений. В основе Системы лежит сервер приложений. Главным языковым средством для разработки приложений в Системе на настоящее время является программное средство Django, которое написано на языке Python. Архитектура сервера приложений Системы (рис. 1) отражает требования к ее функциональным возможностям.

Основу Системы составляет ядро, реализующее базовые (системообразующие) функции, которые напрямую определяют и качество (показатели качества) Системы в целом. Ядро обеспечивает единый интерфейс и базовый каркас добавления, просмотра, редактирования и удаления результатов научно-педагогической деятельности сотрудников. В ядре реализуются функции поиска похожих объектов, которые используются для подбора, в частности, похожих сотрудников, журналов и статей. Важной частью ядра является механизм кэширования. Он используется для ускорения доступа к различным данным Системы. В ядре реализуются также и другие общие функции, например, поддерживающие политику безопасности Системы, в частности, механизмы разграничения доступа к различным категориям данных Системы.

На следующем уровне архитектуры расположены два базовых приложения, модуль организаций и модуль сотрудников. Отметим, что под приложением понимается логически и физически отдельный программный модуль, реализующий ту или иную базовую или прикладную функцию. При этом каждое приложение одновременно является и приложением в терминологии Django («Django app»). Модуль организаций включает классы и логику действий, связанную с данными на уровне организаций, в частности, механизм управления информацией о сотрудниках, которым делегирована роль ответственных от организации и отдельных ее подразделений за сопровождение информации. Приложение «сотрудники» содержит базовые классы, связанные с сотрудниками, в частности, их профиль в Системе и список работ. Модули организаций и сотрудников являются базовыми приложениями потому, что с ними связаны все остальные приложения Системы, которые находятся выше по иерархии.

Основным уровнем архитектуры являются приложения класса «результаты деятельности». Каждое приложение отвечает за отдельный тип результатов научно-инновационной и преподавательской деятельности сотрудников (например, публикации) или за общую сущность, связанную с таким типом (например, журналы). Данный подход позволяет сравнительно легко добавлять новые типы результатов деятельности персонала в Систему. Такие функциональные возможности как добавление, просмотр, редактирование и удаление, а также привязка результата деятельности к сотрудникам-авторам, уже реализованы в ядре. Программисту требуется лишь настроить Систему на конкретный тип результатов деятельности, указав

специфицирующие его свойства. Отметим, что наиболее сложным приложением из рассматриваемого класса являются публикации. Оно использует модуль разбора библиографических ссылок в текстовом формате Freecite который был усовершенствован с целью предоставления пользователю возможности обработки текстов на русском языке.

На следующем уровне иерархии архитектуры Системы расположены два модуля, отвечающие за анализ данных, накопленных на предыдущих уровнях. Модуль статистики содержит функции преимущественно количественного анализа данных Системы. Он включает в себя функции анализа данных на уровне организации, подразделения и отдельного сотрудника. В этом модуле также реализован простой тематический анализ показателей результативности сотрудников. Приложение «статистика» включает функции получения и обработки показателей цитирования отдельных статей из Web of Science, из Scopus, а также поиск статей в этих системах. Кроме этого, в модуль статистики входят механизмы расчета принятых в подразделениях организации коэффициентов эффективности деятельности сотрудников на основе результатов их деятельности. Эти коэффициенты в дальнейшем могут использоваться при расчете поощрительных надбавок сотрудникам.

Приложение «онтологии» имеет целью выполнение более глубокого тематического анализа данных в Системе. В частности, в механизмах этого модуля используются модели предметных областей, построенные в автоматизированном режиме. Для их построения используются анонсы научных конференций, а также результаты запросов к поисковым системам, таким как Bing. Эти модели содержат термины, характерные для соответствующей предметной области. По этим терминам затем выполняется классификация результатов деятельности сотрудников (например, статей), а также подбор похожих объектов для удобной навигации по Системе. Отметим, что большинство функций этого модуля пока находятся в стадии исследовательской разработки. Однако эти исследования рассматриваются разработчиками Системы, как одно из самых перспективных направлений ее развития.

Система ориентирована на эффективное сочетание с одной стороны - интересов организации, в которой она эксплуатируется, и сотрудников её административно-управленческого аппарата, с другой - с интересов отдельных ученых и педагогов. Приоритет в процессах сбора и верификации данных отдается потокам, которые мы именуем потокам «снизу-вверх», от отдельных персоналий в базу данных системы. В этой связи – одно из главных требований к Системе - предельно удобный, дружественный интерфейс для взаимодействия с конечным пользователем и эффективные разноплановые механизмы верификации (очистки) данных. В первую очередь, при этом принимается во внимание тот факт, что только сами ученые могут точно описать и сверить с оригиналом, если это необходимо, свои научные результаты. Здесь следует отметить, что этот принцип не исключает



возможности использования данных, которые могут быть экспортированы, в том числе – в массовом порядке, из других источников с последующей их верификацией. Однако такой способ пополнения Системы данными является вспомогательным.

Наличие большого массива данных в Системе и возможности их разноплановой обработки заставляют нас активно работать над созданием эффективной процедуры оценки достоверности представленных в ней данных. Такая процедура включает иерархически организованный институт ответственных от подразделений МГУ за сопровождение информации в Системе, а также механизмы разграничения доступа к данным разного уровня конфиденциальности. Институт ответственных с четким обозначением сферы их компетенции в работе с данными является важным звеном, реализующим, наряду с программными механизмами, меры административного характера, способствующие поддержанию высокого уровня достоверности данных в Системе. На направлении разграничения доступа к данным задействованы современные модели и программные механизмы, в том числе новые, авторские, разработанные в МГУ.

В настоящее время научно-инновационная педагогическая деятельность ученых характеризуется более чем 20 параметрами, в числе которых: публикации в научных журналах, сборниках статей и тезисы докладов; патенты и свидетельства на интеллектуальный продукт, курсы лекций и семинары; руководство дипломными работами и диссертациями.

Другой важный принцип, которым разработчики руководствуются в процессе развития системы, это высокий уровень открытости информации. Все сведения о результатах научной работы сотрудников доступны в сети Интернет. Для каждого сотрудника университета автоматически формируется его научный профиль, который отражает всю научно-педагогическую деятельность сотрудника.

На рис. 2 в качестве примера представлен научный профиль академика В.П. Скулачева. Такой профиль автоматически формируется по результатам анализа для каждого из пользователей Системы. В нем отражаются сведения, достаточно полно характеризующие результаты его научно-педагогической и инновационной деятельности. На наш взгляд, такой уровень открытости информации не только повышает статус ученого в научном мире, но и способствует повышению ее достоверности, что очень важно в современных условиях.

Система активно используется для оценки степени востребованности результатов ученых МГУ имени М.В.Ломоносова в мире. Главным критерием оценки в данном случае являются индексы цитирования Web of Science и Scopus (рис. 3). Безусловно, они не отражают все аспекты научной деятельности ученого, особенно в гуманитарных областях. Однако индексы цитирования являются основой для многих международных рейтингов учебных заведений и мы вынуждены ориентироваться на эти показатели.



# ИСТИНА

Интеллектуальная Система Тематического Исследования НАучно-технической информации

Главная Моя страница Добавить работу Поиск Статистика О проекте Помощь



## Скулачев Владимир Петрович

VPSkulachev

журнал действит авторизоваться сбросить кэш история изменения объектов отправить сообщение

МГУ имени М.В. Ломоносова, Факультет биоинженерии и биоинформатики, с 1 сентября 2002

МГУ имени М.В. Ломоносова, Научно-исследовательский институт физико-химической биологии им. А.Н.Белозерского, с 14 мая 1991

доктор биологических наук с 1969 года

академик РАН с 15 декабря 1990 г.

член-корреспондент РАН с 26 ноября 1974 г.

профессор по кафедре биоинженерии с 1 июня 2003 г.

Соавторы: Черняк Б.В., Бакеева Л.Е., Семенов А.Ю., Зоров Д.Б., Драчев Л.А., Мохова Е.Н., Антоненко Ю.Н., Констан А.А., Самуилов В.Д., Плетюшкина О.Ю., Плотников Е.Ю., Kaulen A.D., Высоких М.Ю. показать полностью...

281 статья, 4 книги, 2 доклада на конференциях, 7 тезисов докладов, 12 патентов, 6 диссертаций, 3 учебных курса

Количество цитирований статей в журналах по данным Web of Science: 5646, Scopus: 3314

ResearcherID: Vladimir Skulachev

### Информация

Отчеты сотрудника  
Список отобранных статей  
Экспорт публикаций в BibTeX  
Редактировать место работы  
Редактировать степени и звания

### Деятельность

BibTeX

стиль: **обычный** | GOCT | plain | abbrev | acm | alpha | amsalpha | amsplain | apalike | ieee | siam

### СТАТЬИ В ЖУРНАЛАХ

★ 2014 A short-chain alkyl derivative of Rhodamine 19 acts as a mild uncoupler of mitochondria and a neuroprotector  
Khalilova LS, Silachev DN, Rokitskaya TI, Avetisyan AV, Lyamzaev KG, Severina II, Ilyasova TM, Gulyaev MV, Dedukhova VI, Trendeleva TA, Plotnikov EY, Zuyagliskaya RA, Chernyak BV, Zorov DB, Antonenko YN, Skulachev VP  
в журнале *Biochimica et Biophysica Acta - Bioenergetics*, издательство *Elsevier BV (Netherlands)*, том 1837, № 10, с. 1739-1747 DOI

2

Рис. 2. Профиль ученого в системе ИСТИНА.

## Информация о публикациях

Инструкция и правила расчета коэффициентов

Выберите параметры запроса

Тип публикации: Все журналы Только проверенные Метрика: Число ссылок WoS (на статью) Год: Все

Обновить

критерий: число цитирований статей

### Все подразд.

	Name	Value
1	Химический факультет	27586
2	Физический факультет	19302
3	Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына	12833
4	Научно-исследовательский институт физико-химической биологии им. А.Н.Белозерского	8866
5	Факультет наук о материалах	7581
6	Биологический факультет	4472
7	Научно-исследовательский институт механики	3761
8	Механико-математический факультет	2441
9	Государственный астрономический институт имени П.К.Штернберга	2349
10	Геологический факультет	1622

самые «успешные» подразделения

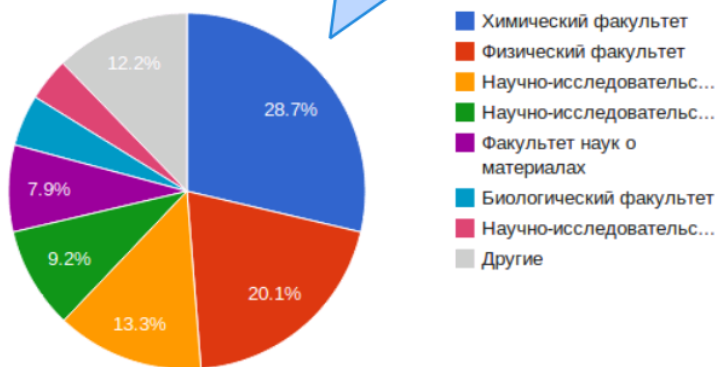


Рис. 3. Рейтинг подразделений МГУ по числу цитирований статей.

Данные о более чем 12,5 тысячах ученых и преподавателей университета, об их научно-инновационной и педагогической деятельности, 353283 статей, 80472 докладов на конференциях, 34668 книг, сведения о 5745 научно-исследовательских работ - далеко не полный перечень, характеризующий объем данных в Системе. Это массив, результаты анализа которого и являются в настоящее время исходными для принятия административно-управляющих решений в МГУ. В сфере кадровой политики – это анкеты из системы «ИСТИНА» на конкурсный отбор, в свете поощрения за результаты научных исследований – данные о публикациях (их рейтинговая ценность), описание наиболее ярких достижений и другие. В конце прошлого года в МГУ оперативно (за 3-4 недели) были проведены 3 конкурса, которые были призваны по итогам года определить ученых и преподавателей, внесших наиболее весомый вклад в достижения МГУ. Исходными для Конкурсной комиссии были результаты сравнительного анализа данных о научно-инновационной и педагогической деятельности конкурсантов, представленных в Системе.

В университете разработан и запущен в эксплуатацию на основе системы «ИСТИНА» Официальный сайт 104 диссертационных советов МГУ. В нем аккумулированы и эффективно обрабатываются все соответствующие Положениям Минобрнауки и ВАК РФ данные по прохождению дел в диссертационных советах МГУ.

Годовая отчетная компания по научно исследовательской работе уже второй год проведена на основе анализа данных в системе «ИСТИНА».

Уже сегодня мы завершаем работу по переходу на новые принципы конкурсного избрания на научные должности в МГУ. В их основе механизмы Системы «ИСТИНА», которые позволяют каждому ученому и руководителю подразделения оперативно посчитать:

- персональный рейтинг любого научного сотрудника подразделения;
- на основе агрегирования и определения (по медиане) этих рейтингов получить средние по подразделению рейтинги для отдельных категорий сотрудников (м.н.с., н.с., в.н.с., и гл.н.с.);
- гистограмму – зависимость числа сотрудников, имеющих тот или иной рейтинговый балл, от величины этого балла.

С этой целью в каждом из структурных подразделений МГУ (факультеты, институты, центры) разработана формула расчета показателя эффективности работы отдельного научного работника (рис. 4).

Категория	Ограничения [?]	Свойства	Вес	Действия
Статьи (1.1.1) Статьи → Журналы → В России → ВАК		Импакт-фактор > РИНЦ	60	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Зарубежные статьи (1.1.2) Статьи → Журналы → Международные		Импакт-фактор > ISI	30	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Тезисы международные (1.2.1) Статьи → Тезисы конференции → Международные		Количество	10	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Тезисы российские (1.2.2) Статьи → Тезисы конференции → В России		Количество	5	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Патенты (II) Патенты		Количество	20	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Участие в конференциях (IV) Участие в конференции → Международные		Количество	6	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Участие в конференциях (Россия) (IV) Участие в конференции → В России		Количество	4	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Канд. диссертация (V) Защищенные диссертации → Кандидатская		Количество	30	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Докт. диссертация (V) Защищенные диссертации → Докторская		Количество	60	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Учебная работа (VI) Учебная работа → Курс лекций → ½ года		Количество	15	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>
Учебная работа (VI) Учебная работа → Курс лекций → 1 год		Количество	30	<input type="button" value="↶"/> <input type="button" value="↷"/> <input type="button" value="↓"/> <input type="button" value="X"/>

Рис. 4. Формирование формулы расчета показателя эффективности работы научных работников в системе ИСТИНА.

Для оценки эффективности работы научных сотрудников в рамках Системы построен "Конструктор формул расчета эффективности работы сотрудника". Этот конструктор позволяет ответственным пользователям Системы сформировать необходимую формулу расчета, которая может учитывать практически все виды работ, внесенных в Систему. Конструктор обладает достаточной гибкостью для адекватного учета специфики того или иного подразделения (факультета, института, центра). Подготовленная с помощью такого конструктора формула расчета эффективности может использоваться как для построения таблицы с указанием полученных баллов для всех сотрудников подразделения (рис. 5), так и для составления персонального отчета (рис. 6), который содержит перечень всех работ конкретного сотрудника, с указанием полученного числа баллов за каждую работу. Персональные отчеты позволяют проверить достоверность интегральных показателей эффективности.

Даты работ: начало 2010 | окончание 2014 | Посчитать средние значения

Научно-исследовательский институт механики

Применить | Экспорт в CSV (Excel)

Show 5 entries | Search:

ID	Сотрудник	Подразделение	Должность	Сумма	Статьи (1.1.1)	Зарубежные статьи (1.1.2)	Тезисы международные (1.2.1)	Тезисы российские (1.2.2)	Патенты (II)
11625	Козицын Александр Сергеевич	404 Лаборатория автоматизации экспериментальных исследований		529.8	37.8	0	0	30	0
11626	Афонин Сергей Александрович	404 Лаборатория автоматизации экспериментальных исследований		493.08	109.08	12	0	30	0
11628	Коршунов Андрей Александрович	404 Лаборатория автоматизации экспериментальных исследований		178.84	12.84	0	10	25	0
11633	Голомазов Денис Дмитриевич	404 Лаборатория автоматизации экспериментальных исследований		221.4	83.4	0	0	35	0
11634	Васенин Валерий Александрович	404 Лаборатория автоматизации экспериментальных исследований		1459.523	339.6	13.98	30	40	0

Showing 1 to 5 of 207 entries | Previous 1 2 3 4 5 ... 42 Next

Рис. 5. Таблица баллов сотрудников подразделения.

Результаты расчёта по формуле "Формула НИИ механики МГУ" с диапазоном дат 2012-2014

Васенин Валерий Александрович, общая сумма баллов: **912,2** [просмотреть средние значения на сайте МГУ](#)

Название работы	Тип [?]	Вес [?]	Множитель [?]	Авторов [?]	Баллы [?]
Статья (1.1.1)					<b>224,28</b>
Васенин В.А., Кривчиков М.А., Статическая семантика стандарта ЕСМА-335. Программирование 4, 3-16, 2012	ВАК	60	0,38	2	23,1
Васенин В.А., Афонин С.А., Козицын А.С., Голомазов Д.Д., Бахтин А.В., Ганкин Г.М., Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА) Обзорные прикладной и промышленной математики (19)2, 239-240, 2012	ВАК	60	0,03	6	2,28
Васенин В.А., К созданию международной системы мониторинга и анализа информационного пространства для предотвращения и прекращения военно-политических киберконфликтов. Информационные технологии 9, 2-10, 2012	ВАК	60	0,33	1	19,92
Васенин В.А., Модернизация экономики и новые аспекты инженерии программ. Программная инженерия 2, 2-17, 2012	ВАК	60	0,21	1	12,84
Васенин В.А., Кривчиков М.А., Крошилин А.Е., Крошилин В.Е., Рагулин А.Д., Роганов В.А., Распараллеливание расчетного кода улучшенной оценки "БАГИРА" для моделирования трехмерной теплогидродинамики многофазных сред в составе полномасштабной суперкомпьютерной модели "Виртуальная АЭС. Программная инженерия 6, 15-23, 2012	ВАК	60	0,21	6	12,84
Васенин В.А., Итхес А.А., Пучков Ф.М., Шалченко К.А., Обеспечение информационной безопасности в распределенных системах на основе технологий Grid и Cloud Computing, традиционные средства защиты и вопросы асимметрии доверия. Программная инженерия 1, 28-34, 2012	ВАК	60	0,21	4	12,84
Васенин В.А., Левин В.Ю., Пучков Ф.М., Шалченко К.А., К созданию защищенной наложенной сети связи нового поколения для передачи данных. Программная инженерия 4, 2-8, 2012	ВАК	60	0,21	4	12,84
Васенин В.А., Шундеев А.С., Эволюция технологии грид. Информационные технологии 1, 2-9, 2012	ВАК	60	0,33	2	19,92
Васенин В.А., Афонин С.А., Голомазов Д.Д., Козицын А.С., Интеллектуальная Система Тематического Исследования Научно-технической информации (ИСТИНА). Информационное общество 1-2, 21-36, 2013	ВАК	60	0,34	4	20,4
Васенин В.А., Кривчиков М.А., Модель динамического параллельного исполнения программ. Программирование 1, 45-59, 2013	ВАК	60	0,38	2	23,1
Васенин В.А., Афонин С.А., Панюшкин Д.С., Модели распространения информации в социальных сетях: тестовые испытания. Программная инженерия 4, 35-41, 2014	ВАК	60	0,21	3	12,84

Рис. 6. Персональный отчет сотрудника в системе ИСТИНА с указанием баллов.

Как следствие, появляется возможность у каждого сотрудника, также как и у ответственного за сопровождение информации от подразделения, в котором он работает, посчитать или увидеть на своей странице в ИСТИНЕ индивидуальный рейтинг, коррелирующий с данными его персональной страницы. Как следствие, каждый ученый может определиться с тем, как его повысить, если он ниже пороговых значений, которые он тоже может увидеть в ИАС «ИСТИНА». Здесь следует отметить, что с помощью программных механизмов логического разграничения доступа данные по персональному рейтингу уже являются конфиденциальными. Они доступны только пользователю – персоналию и научно-административному персоналу, стоящему выше по управленческой иерархии организации (университета). За два года развития Системы накоплен значительный опыт. Многого уже удалось добиться, однако большое число задач, в том числе теоретического характера и прикладных, еще предстоит решить. Уже сегодня мы готовы предоставлять доступ ко всем функциям системы другим университетам и научным центрам России, создавая аналоги системы «Наука МГУ» - «ИСТИНА» в интересах этих организаций. Надеемся, что на таком подходе, достаточно быстро можно будет создать исходную базу для Карты науки России, которая будет адекватно отражать состояние дел с научно-инновационной, а в ближайшей перспективе и - образовательной деятельностью в стране.

## ЛИТЕРАТУРА

1. Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА) / С.А. Афонин и др. Под ред. академика В.А.Садовниченко. – М.: Издательство Московского университета, 2014. – 262 с.

# ПОСТРОЕНИЕ СРЕДЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ НАНОКОМПОЗИТНЫХ МАТЕРИАЛОВ НА БАЗЕ КОМПЛЕКСА ВЫЧИСЛИТЕЛЬНЫХ СЕРВИСОВ И ВЫСОКОУРОВНЕВЫХ ВЕБ-ИНТЕРФЕЙСОВ

В.М. Волохов<sup>1</sup>, Д.А. Варламов<sup>1,2</sup>, А.В. Волохов<sup>1</sup>, А.И. Прохоров<sup>1</sup>, Г.А. Покатович<sup>1</sup>

<sup>1</sup> *Институт проблем химической физики РАН, г. Черноголовка, РФ*

<sup>2</sup> *Институт экспериментальной минералогии РАН, г. Черноголовка, РФ*

Современное мировое состояние вычислительной химии характеризуется тотальным использованием мощных вычислительных ресурсов как параллельных, так и распределенных, для решения задач различных классов. В настоящее время в крупнейших мировых научных центрах, специализирующихся в теоретической физике и теории химических превращений, ведутся активные работы по разработке наиболее эффективных вычислительных процедур для параллельных, «гибридных» и распределенных вычислительных сред, позволяющих исследовать самые разнообразные химические процессы. Потенциально задачи в области молекулярного моделирования и многоуровневого иерархического моделирования материальных объектов (от квантово-механического уровня до уровня сплошных сред и конструкций) могут выходить на уровень востребованности многих петафлопс. Некоторые задачи оптимизации крупных молекулярных структур требуют выполнения до  $10^9$  отдельных расчетов. Типичный докинг белковых лигандов с размерностью  $200 \text{ атомов} \times 300 \text{ 000 конфигураций} \times 1000 \text{ CPU-часов}$  требует до 300 Пф вычислительных мощностей. Для построения многомерных потенциальных поверхностей, адекватно описывающих химические реакции, нужно провести  $10^2$ – $10^N$  независимых весьма ресурсоемких *ab initio* расчетов. Детальное моделирование даже отдельных этапов каталитических реакций в топливных элементах на основе водорода требует до 60000 CPU-часов. Для исследования динамики не очень сложной химической реакции с использованием метода классических траекторий зачастую нужно рассчитать до  $10^7$ – $10^9$  независимых траекторий. Численное исследование многопараметрических функций  $F(x_1, x_2 \dots x_n)$  в области изменения параметров  $x_1, x_2 \dots x_n$  при разбиении диапазона изменения каждого параметра на 10 ячеек требует проведения  $10_N$  независимых расчетов  $F$  – и так далее. При этом

Опыт работ авторов в области вычислительной химии [1-3], в том числе при моделировании наноконструктов и наноструктурированных объектов [4, 5] показывает, что, помимо использования суперкомпьютерных центров класса «Ломоносов» или МВС-10000, весьма перспективным для целей компьютерного моделирования вещества является создание проблемно-ориентированных ресурсов среднего размера, направленных на решение

определенных классов задач, при этом позволяющих: 1) охватить все стадии моделирования (нано)объектов от формирования исходных данных и выбора методов расчетов до получения интерактивных компьютерных моделей и баз данных результатов моделирования; 2) предоставить доступ к средствам моделирования максимально широкому кругу пользователей через организацию максимально дружелюбных пользователю веб-сервисов; 3) вовлекать в расчеты все доступные вычислительные ресурсы – локальные кластеры, «гибридные» установки, ресурсные сайты распределенных сред; 4) оказывать интеллектуальную поддержку пользователю при выборе стратегии моделирования и использования методов расчетов; 5) объединять в рамках единого комплекса прикладные пакеты, ориентированные на разные аспекты моделирования (нано)объектов – создание «вычислительных последовательностей», при этом выбирая наиболее оптимальные варианты установки данных пакетов и наиболее подходящие пулы ресурсов для них.

При этом может быть во многом решен целый ряд проблем, характерных для квантово-химических расчетов и молекулярного моделирования и обычно существенно ограничивающих работу исследователей:

- подавляющее большинство прикладных пакетов квантовой химии и молекулярной динамики в качестве исходных данных требует введения описания до нескольких сот параметров (в том числе весьма неявно связанных между собой) и методов расчетов, многих сотен и даже тысяч первичных координат атомов, полевых потенциалов, сложной иерархии используемых констант, данных и вычислительных процедур, многовариантности построения моделей и т.п. Их подготовка очень трудоемка, требует детального знания синтаксиса входных файлов, тщательной проверки введенных данных и практически затруднена для пользователя (часто обнаружение ошибки несогласованности данных обнаруживаются на этапе непосредственно выполнения расчетов, при уже большом затраченном времени на выполнение расчета);
- для построения и хранения мультипараметрических моделей требуется создание и поддержка достаточно объемных структурированных баз данных. На основе хранимых данных будет возможно создавать «шаблоны» последующих проектов;
- большие временные затраты на расчет моделей (например, до 40-50 часов на расчетный шаг на СК «Ломоносов» при необходимых 150-200 шагах для расчета транспорта единичного протона через наномембрану) требуют применения на ресурсах новейших методов проведения расчетов, основанных на использовании «гибридных» и распределенных моделей вычислений;
- прикладные пакеты в результате вычислений создают огромные (вплоть до многих гигабайтов) «сырые» (raw) файлы результатов, требующие постпроцессинга данных, генерализации, упрощения и перевода в другие форматы для наглядной интерактивной визуализации и сопряжения с



другими прикладными пакетами. Как правило, на основе результатов требуется построение 3-D моделей молекулярных структур, позволяющих проводить различные манипуляции с ними и получать наиболее эффективное представление о данных структурах. Кроме того, требуется формировать «на лету» большое количество текстово-табличной и графической информации (бинарные и более сложные диаграммы), позволяющей адекватно оценивать качество созданных моделей в зависимости от входных параметров.

Большинство этих проблем плохо решаемы непосредственно в зоне ответственности суперкомпьютерных центров ввиду «всеядности» последних, поэтому необходимо построение «внешних» проблемно-ориентированных «шлюзов». В области собственно химии (исключая биохимию и фармацевтику) сейчас почти отсутствуют комплексные решения, позволяющие снять эти проблемы, поэтому разработка современного комплексного компьютерного инструментария для максимального упрощения работы с квантово-химическими и молекулярно-динамическими прикладными пакетами вкупе со значительной интенсификацией расчетов на основе суперкомпьютеров и распределенных сред способна коренным образом улучшить ситуацию в области молекулярного моделирования нанокompозитных материалов и нанобъектов на их основе. Создание проблемно-ориентированных вычислительных RIA-сервисов с дистанционным доступом пользователей является одним из основных трендов развития прикладных вычислений в разных областях науки и должно стать эффективным научным инструментом для российских исследователей. Развитие же программных средств создания высокоэффективных веб- и грид-сервисов в настоящее время позволяет создавать инструменты, контролирующие все стадии процесса моделирования: от ввода первичных данных до интерактивной 3-D визуализации объектов или процессов. Подобные программно-аппаратные комплексы позволят решать на принципиально новом вычислительном уровне многочисленные «тяжелые» задачи в следующих областях: строение молекул и структур твердых тел, кинетика и механизмы сложных химических реакций, описание процессов образования и модификации полимеров, биологических процессов и систем, а также в области общих проблем химической физики.

Опыт участия авторов в течение 2004-2013 годов в ряде российских грид-проектов [2, 3], как-то: EGEE-RDIG и EGI-RU-NGI, СКИФ-Грид, ГридННС («Национальная нанотехнологическая Сеть»), «Развитие пилотной зоны российской грид-системы для высокопроизводительных вычислений...», различных программ Президиума РАН, ориентированных в том числе на вовлечение проблемно-ориентированных прикладных сервисов в грид-вычисления, позволила сформировать оптимальную модель прикладной компьютерной среды (программно-аппаратного комплекса), направленной (в нашем случае) на эффективное проведение широкомасштабных вычислений в

области моделирования наноматериалов через веб-средства дистанционного доступа (рис.1)

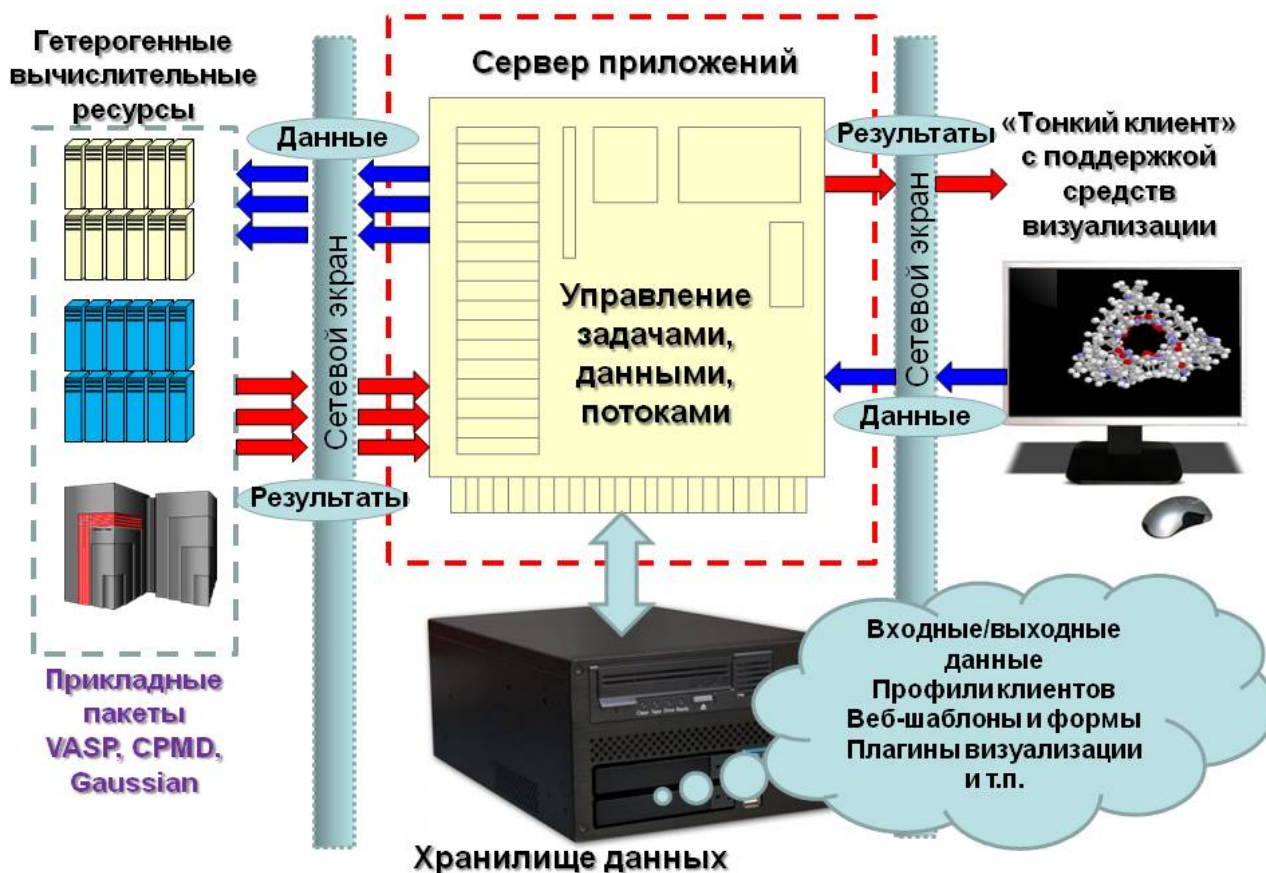


Рис. 1. Структура среды компьютерного моделирования нанокompозитных материалов на базе высокопроизводительных вычислительных сервисов и высокоуровневых веб-интерфейсов к ним.

Подобный комплекс способен работать как в «одиночном» режиме (с вовлечением локальных ресурсов – кластеров, «гибридных» GPU-установок, мультядерных высокопроизводительных станций), так и с удаленными ресурсами (при условии соблюдения достаточных мер безопасности) и (при установке промежуточного Grid-middleware – шлюзы, ресурсные грид-сайты, грид-хранилища) в составе грид-полигонов. При этом доступ к ресурсам, а также управление комплексом осуществляется через высокоуровневые, дружелюбные конечному пользователю веб-интерфейсы. Среда объединяет вычислительные сервисы, созданные на основе прикладных квантово-химических и молекулярно-динамических пакетов (пакеты VASP, CPMD), в том числе с применением «гибридных» технологий расчетов на основе GPU- и мультядерных ускорителей; интернет-ориентированный инструментарий для пользовательского доступа к этим сервисам в виде комплекса высокоуровневых веб-интерфейсов; средства размещения, хранения и обработки первичных

данных и полученных результатов. В состав высокоуровневых веб-интерфейсов помимо стандартных приемов дистанционной работы с прикладными пакетами вводятся дружелюбные пользователю средства по формированию и тестированию сложных конфигурационных файлов и массивов входящих данных, инструменты анализа и предварительной обработки полученных результатов. Также доступны средства запуска и мониторинга выполняемых заданий.

В настоящее время в ИПХФ РАН реализован пилотный вариант подобной системы, основанный на локальных ресурсах (кластер на 176 двухпроцессорных узлов HP Proliant производительностью до 15 Тфлопс и «гибридная» установка на базе GPU Tesla C2075 – 2,6 Тфлопс), с поддержкой прикладных пакетов VASP (Vienna Ab initio Simulation Package, <https://www.vasp.at>) и CPMD (Car-Parrinello Molecular Dynamics, <http://cpmd.org>). Система иерархических веб-интерфейсов позволяет авторизовать пользователя комплекса, проводить мониторинг доступных ресурсов, подготовить проект вычислительного задания (включая создание и редакцию начальных данных и конфигурационных файлов, в том числе на основе имеющихся шаблонов); запустить прикладной пакет в инфраструктуре комплекса на выбранном ресурсе; вести мониторинг выполнения задания (включая останов и перезапуск); по завершении – получить результаты счета, провести их пост-обработку (генерализацию, архивацию, конвертацию в другие форматы) и отправку в хранилище, а также обеспечить визуализацию смоделированных объектов и структур.

Основным подходом к решению поставленной задачи является комбинированное использование большинства современных вычислительных технологий для компьютерного моделирования в области квантовой химии и молекулярной динамики в рамках единой компьютерной среды дружественной конечному пользователю-химику, позволяющей эффективно проводить расчеты в удаленном режиме. Создаваемая вычислительная среда позволяет более эффективно решать "тяжелые", многостадийные задачи в областях молекулярного моделирования наноматериалов и наноструктур, делая их доступными более широкому кругу исследователей. Модульная структура комплекса (вычислительной среды) позволяет достаточно просто вводить новые методы расчетов, выбирать различные оптимальные варианты сочетания вычислительных сервисов, вводить новые функциональные возможности (в том числе новые прикладные пакеты) без ущерба общей работоспособности комплекса.

С использованием комплекса проведено более 50 вычислительных экспериментов в области компьютерного моделирования компонентов низкотемпературных топливных элементов [1, 4] (нанокompозитные электроды на основе диоксидов переходных металлов и платины и мембраны на основе сложных гетерополикислот) и литий-ионных источников тока новых типов [5]

(композитные структуры – нанотрубки, наностержни, структурные кластеры кремния размерностью до 300 атомов на основе углерода и кремния).

Проведение работ по данной тематике поддержано грантом РФФИ № 15-07-07867-а, рук. Волохов В.М.).

## ЛИТЕРАТУРА

1. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, Г.А. Покатович, А.В. Волохов. Квантово-химическое моделирование процессов в низкотемпературных электрохимических топливных элементах // Альманах «Суперкомпьютерные технологии в науке, образовании и промышленности», (под ред. акад. В.А.Садовниченко, акад. Г.И.Савина, чл.-корр. Вл.В.Воеводина). – М., изд-во Московского университета, 2013, вып.5, 208 с., с.172-176
2. В.М.Волохов, Д.А.Варламов, А.В.Волохов, А.В.Пивушков, Г.А.Покатович, Н.Ф.Сурков. Грид-сервисы в вычислительной химии: достижения и перспективы // "Вестник Уфимского Государственного авиационно-технического университета. Серия «Управление, вычислительная техника и информатика»", 2011, т. 15, № 5 (45), с.161-169
3. Пивушков А.В., Волохов В.М., Варламов Д.А., Волохов А.В. Применение новых вычислительных технологий для повышения эффективности расчетов в грид-средах // "Вестник Уфимского Государственного авиационно-технического университета. Серия «Управление, вычислительная техника и информатика»", 2013, т.17, № 2 (55), с.117-124
4. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, А.В. Волохов, Г.А. Покатович. Квантово-химическое моделирование наноструктурированных компонентов низкотемпературных электрохимических топливных элементов // "Параллельные вычислительные технологии (ПаВТ'2014)", труды международной научной конференции (1–3 апреля 2014 г., г. Ростов-на-Дону). Челябинск: Издательский центр ЮУрГУ, 2014, (389 с.), с.235-242
5. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, А.В. Волохов, Г.А. Покатович. Компьютерное моделирование структуры и свойств наноматериалов и наноструктурированных электродов в новых типах Li-ионных источников тока // "Параллельные вычислительные технологии (ПаВТ'2015)": труды международной научной конференции (31 марта – 2 апреля 2015 г., г. Екатеринбург). Челябинск: Издательский центр ЮУрГУ, 2015. (536 с.), с.370-376

# КОЛИЧЕСТВЕННАЯ ОЦЕНКА МЕТАКОМПЕТЕНЦИЙ УЧАЩЕГОСЯ НА ОСНОВЕ АНАЛИЗА ЕГО ПОВЕДЕНИЯ В СОЦИАЛЬНЫХ СЕТЯХ

Е.А. Гаврилина, М.А. Захаров, А.П. Карпенко, Е.В. Смирнова

*МГТУ им. Н.Э. Баумана*

## Введение

Различаем метапредметные, метакреативные и метакогнитивные (мета) компетенции учащегося [1, 2]. Список известных работ, посвященных различным аспектам формирования и оценки метакомпетенций учащихся, невелик. Основные из этих работ рассмотрены нами в публикации [2].

Вообще говоря, метакомпетенции учащегося могут быть оценены следующим образом: на основе подготовленных им документов; путем анализа поведения учащегося в системах управления ВУЗом, различных системах тестирования, системах управления обучением *LMS*, массовых открытых онлайн курсах *МООС*; с использованием данных полученных из социальных сетей, участниками которых является учащийся.

Работа посвящена оценке метакомпетенций учащихся на основе анализа их поведения в таких социальных сетях, как *Twitter*, *Facebook*, *Vk*, *Odnoklassniki*, *Linkedin*. Точнее говоря, речь идет об оценке метакомпетенций на основе анализа прямых (личные данные, высказывания, комментарии) и косвенных данных (подписанные группы, события, места, другие участники), извлеченных из указанных социальных сетей. Разрабатываемые методы извлечения данных из указанных социальных сетей и методы оценки метакомпетенций реализуются в рамках программной системы *META*<sup>3</sup>, предназначенной для использования в учебных заведениях и кадровых агентствах компаний.

## 1. Индикаторы метакомпетенций

*Метапредметные* результаты учащегося рассматриваем как освоенные им метапредметные понятия и так называемые универсальные учебные действия (регулятивные, познавательные, коммуникативные). Другими словами, полагаем, что составляющими метапредметности являются освоенные метапонятия, а также регулятивные, познавательные, коммуникативные компетенции. Количественные оценки указанных составляющих называем *индикаторами метапредметности*.

*Метакогнитивные* процессы были выделены Дж. Флэвеллом, который определил их как индивидуальное знание, касающееся собственных когнитивных процессов и результатов познавательной деятельности. Выделяем *индикаторы метакогнитивности* учащегося, оценивающие его абстрактное

мышление, вербальные способности, математические способности, перцептивные способности, пространственное мышление, техническое мышление.

*Метакреативность.* Креативность в наиболее общем виде понимается как способность человека выходить за рамки стимульной ситуации, предлагать ответы, отличающиеся высокой оригинальностью и нестандартностью мышления. Метакреативность можно определить как интегральное качество личности, обеспечивающее не только возможность выхода личности за рамки стимульной ситуации, но и способность осознавать, как это выход осуществляется, выбирать наиболее адекватные стратегии для него. Предлагаем индикаторы метакреативности учащегося, основанные на оценке гибкости его мышления, продуктивности мышления, беглости мышления, оригинальности мышления, разработанности проблемы.

Для всех рассматриваемых типов метакомпетенций выделяем следующие уровни их освоения: декларативное знание; концептуальное знание; процедурное знание; ситуационное знание; поведенческое знание [2].

В силу ограниченности размера статьи в пунктах 2, 3 рассматриваем только те индикаторы, которые определяют метапредметность учащегося.

## **2. Извлечения метапредметной информации из социальных сетей**

Обычно из социальных сетей извлекают анкетные данные пользователя, информацию о его социальных связях и круге ближайших друзей, векторе интересов, образовании и месте работы, информацию о перемещениях человека, а также специфическую информацию об активности пользователя в социальной сети (число постов, альбомов, стандартное время посещения сети). Методы извлечения контента из социальных сетей хорошо известны. Доступен ряд соответствующих программных продуктов, из которых наиболее известны *SocioHub* [3] и *Crawler* [4].

В контексте оценки метапредметности интерес представляют используемые учащимся понятия и метапонятия, извлекаемые из его контента. Программная система *META*<sup>3</sup> (п. 6) реализует извлечения контента пользователя из всех указанных выше социальных сетей. Для извлечения информации, например, из сети ВКонтакте, используем следующие группы методов: стена (*wall*); виджеты (*widgets*); страницы (*sites*); группы (*groups*); доска обсуждений (*board*); заметки (*comments*); места (*places*); сообщения (*messages*); новости (*newsfeed*); мне нравится (*likes*); опросы (*polls*); документы (*docs*); закладки (*fave*).

Все тексты, содержащиеся в извлеченном из данной социальной сети контенте, подвергаем семантическому анализу. Используем алгоритмы, основанные на правилах, то есть алгоритмы, которые подразумевают наличие заранее составленных словарей, описывающих варианты использования лексических единиц в тексте [5]. Указанные словари строим на основе

онтологий  $\{O_i\}$  соответствующих учебных предметов, формализованных в виде семантических сетей  $\{SS_i\}$ . Индекс  $i$  для простоты записи далее опускаем.

Семантическую сеть онтологии  $O$  определяет кортеж

$$SS = \langle C, R, V, W \rangle, \quad (1)$$

где  $C = \{c_j\}$  - множество понятий сети;  $R = \{r_j\}$  - множество отношений между понятиями набора  $C$ ;  $V = \{v_j\}$ ,  $W = \{w_j\}$  - значения мер важности понятий набора  $C$  и отношений  $R$  соответственно [6]. Методы оценки этих мер рассматриваем п. 4.

Если понятия  $c_j, c_k$ ,  $j \neq k$  в сети  $SS$  связаны между собой некоторым отношением из числа отношений  $R$ , то говорим, что эти понятия связаны информационно. Сопоставляем сети (1) взвешенный ориентированный граф  $G$ , вершины которого соответствуют понятиям онтологии  $O$ , а дуги – информационным связям этих понятий между собой. Веса вершин графа  $G$  равны важностям понятий, а веса дуг – важностям соответствующих отношений.

### 3. Построение когнитивной карты учащегося

Методика контроля понятийных знаний учащегося, основанная на использовании его когнитивной карты, рассмотрена, например, в работе [7]. Когнитивная карта учащегося формализует его представления о некотором фрагменте предметной семантической сети и, в идеале, совпадает с этим фрагментом. Контроль усвоения учащимся понятийного состава данного фрагмента семантической сети сводится к сравнению подграфа сети, соответствующего этому фрагменту, и графа, который определяет соответствующая когнитивная карта учащегося.

Когнитивную карту учащегося, соответствующую семантической сети  $SS$ , определяет кортеж  $CM = \langle \tilde{C}, \tilde{R} \rangle$ , где  $\tilde{C}$  - набор понятий множества  $C$ , которые в карте  $CM$  указаны учащимся как принадлежащие этой сети;  $\tilde{R}$  - набор отношений множества  $R$ , которые в карте связывают понятия набора  $\tilde{C}$  между собой [7].

Специфика извлечения информации из социальных сетей не позволяет получить данные об отношениях  $\tilde{R}$ . Поэтому для оценки метапредметных знаний учащегося используем модифицированные когнитивные карты вида

$$CM = \langle \tilde{C}, \tilde{F}, \tilde{H} \rangle, \quad (2)$$

где  $\tilde{F} = \{\tilde{f}_j\}$ ,  $\tilde{H} = \{\tilde{h}_{j,k}\}$  - веса понятий набора  $\tilde{C}$  и взаимосвязей этих понятий соответственно. Вес понятия  $\tilde{c}_j$  полагаем равным числу использований учащимся этого понятия в анализируемых текстах. Веса  $\{\tilde{h}_{j,k}\}$  определяем на основе *расстояний* между понятиями [7].

Когнитивной карте (2) ставим в соответствие неориентированный, возможно, несвязный взвешенный граф без контуров  $\tilde{G}$ , вершины которого

соответствуют понятиям  $\tilde{C}$ , а дуги – взаимосвязям этих понятий. Веса вершины  $\tilde{c}_j$  и ребра, связывающего вершины  $\tilde{c}_j, \tilde{c}_k$ , равны  $\tilde{f}_j, \tilde{h}_{j,k}$  соответственно.

#### 4. Количественная оценка метапредметности учащегося

**Меры важности понятий и отношений.** Для оценки мер важности понятий набора  $C$  и отношений  $R$  семантической сети  $SS$  используем методы, основанные на кластеризации этой сети [6]. Если понятия  $c_i, c_j$  сети  $SS$  связаны между собой отношением  $r_k \in R$ , то говорим, что эти понятия связаны указанным отношением в узком смысле. Совокупность всех понятий семантической сети  $SS$ , включая понятие  $c_i$ , которые связаны отношением  $r_k$  с понятием  $c_i$  в узком смысле называем  $r_k$ -локальным кластером понятия  $c_i$  и обозначаем  $\omega_k(c_i)$ .

Пусть в узком смысле отношением  $r_k$  понятие  $c_i$  связано с понятием  $c_j$ , которое связано с понятием  $c_k$ , и так далее до понятия  $c_q$ . В таком случае говорим, что понятия  $c_i, c_q$  связаны отношением  $r_k$  в широком смысле. Аналогично локальному определяем  $r_k$ -глобальный кластер  $\Omega_k(c_i)$ .

Часть кластера  $\Omega_k(c_i)$ , которая включает в себя все понятия, расположенные на расстоянии  $\rho = 1, 2, \dots$  от понятия  $c_i$ , обозначаем  $\Omega_k(\rho, c_i)$ . Меры, которые могут быть использованы в качестве расстояния  $\rho$  между понятиями определены, например, в [6]. Взвешенные мультиграфы, соответствующие кластерам  $\omega_k(c_i), \Omega_k(c_i), \Omega_k(\rho, c_i)$  обозначаем  $g_k(c_i), G_k(c_i), G_k(\rho, c_i)$  соответственно.

В качестве важности  $w_k$  отношений набора  $R$  используем общее число понятий семантической сети  $SS$ , связанных между собой отношением  $r_k$ . Если отношение  $r_k$  представляет собой отношение частичного порядка, то в качестве этой величины может быть использована максимальная из высот графов  $G_k(c_i)$ . Важность  $w_k$  определяет также средняя из этих высот. Для произвольного отношения  $r_k$  в качестве величины  $w_k$  используются максимальный и средний диаметры указанных графов. Величину  $w_k$  определяют, наконец, максимальная и средняя реберные плотности графов  $G_k(\rho, c_i)$  [6].

Важность  $v_i$  понятия  $c_i$  можно считать равной взвешенному числу понятий, содержащихся во всех кластерах  $\omega_k(c_i)$ . Можно также использовать взвешенное число понятий, содержащихся во всех кластерах  $\Omega_k(c_i)$ . Если задаться законом изменения влияния понятий на важность данного понятия в зависимости от расстояния до него, то по аналогии определяются важности понятий на основе графов  $G_k(c_i), G_k(\rho, c_i)$ .

**Метрики оценки метапредметности учащегося.** Программная система *МЕТА*<sup>3</sup> (п. 6) предоставляет пользователю возможность самому определить



метрики близости семантической сети  $SS$  и соответствующей когнитивной карты учащегося  $CM$ . Предопределенными в системе являются следующие метрики.

Метрика

$$\mu_1 = \frac{|\tilde{C}|}{|C|}$$

имеет смысл относительного числа понятий набора  $\tilde{C}$ , содержащихся в наборе  $C$ .

Метрика

$$\mu_2 = \frac{\sum_j w_j}{\sum_k \tilde{f}_k}, \quad j \in [1:|C|], k \in [1:|\tilde{C}|]$$

представляет собой аналогичное  $\mu_1$  отношение взвешенных чисел понятий наборов  $\tilde{C}$ ,  $C$ .

Метрика  $\mu_3$  построена на основе близости пар понятий семантической сети  $SS$  в соответствующей когнитивной карте учащегося  $CM$ .

## 5. Используемые методы машинного обучения

Рассматриваемую выборку учащихся обозначаем  $U$  и полагаем, что эта выборка может быть разделена на непересекающиеся обучающую  $U^L$  и тестовую  $U^T$  выборки. Пусть значения индикаторов составляющих данной метакомпетентности тем или иным образом оценены для всех или некоторой части учащихся выборки  $U$  с помощью нормализованной целочисленной шкалы  $Q$ . Набор оценок учащегося  $u_j \in U$  по шкале  $Q$  образует  $|\Phi|$ -мерный вектор  $\Phi_j = (\phi_{j,k} \in Q, k \in [1:|\Phi|])$ . Полагаем, что для каждого учащегося  $u_j$  известна интегральная оценка учителя (отклик)  $e_j = e(u_j) = e(\Phi_j)$ , определенная на основе соответствующих значений индикаторов  $\phi_{j,k}$  на аналогичной  $Q$  целочисленной шкале  $P$ . Каждому набору  $\Phi_j$  оценок учащегося  $u_j$  ставим в соответствие точку в  $|\Phi|$ -мерном *индикаторном* пространстве с координатами  $\phi_{j,k}$ . Важно, что поскольку множество  $\{\Phi_j\}$  всех возможных точек  $\Phi_j$  может иметь большую мощность, получение откликов  $e_j$  для всех точек указанного множества невозможно [2].

*Обучение с учителем (supervised learning)* предполагает, что для каждого объекта выборки  $U$  известен соответствующий отклик (такая выборка называется *размеченной*). Обучение с учителем адекватно задаче *классификации*, то есть алгоритм обучения (стратегия) в данном случае представляет собой алгоритм классификации. Известно большое число алгоритмов классификации и их открытых программных реализаций (см. например, [8]).

*Обучение без учителя (unsupervised learning)* исходит из того, что объекты не размеченной выборки  $U$  требуется сгруппировать в кластеры на основе их близости в индикаторном пространстве. Обучение без учителя адекватно задаче *кластеризации*, то есть стратегия в данном случае есть не что иное, как алгоритм кластеризации. Классические алгоритмы кластеризации означают, по сути, использование кусочно-постоянной аппроксимации функции предпочтений  $e(\Phi)$  учителя в пределах каждого из кластеров. Такая аппроксимация может представлять собой слишком грубую модель учителя, поскольку в силу возможной противоречивости оценок объектов учителем, близость векторов индикаторов  $\Phi_{j_1}, \Phi_{j_2}$  в индикаторном пространстве, вообще говоря, не означает близость откликов  $e_{j_1}, e_{j_2}$ . Поэтому используем нечеткую аппроксимацию его функции предпочтений.

*Обучение с частичным привлечением учителя (Semi-Supervised Learning, SSL)* предполагает, что для части выборки известны объекты и соответствующие отклики (как при обучении с учителем), а для части - только объекты (как при обучении без учителя). В терминах исходной задачи оценки уровня метакомпетенций такая ситуация означает, что для одной части учащихся учителем назначены интегральные оценки, а для другой части по той или иной причине – нет [9]. Известно значительное число методов *SSL*, которые можно разделить на следующие группы: простые эвристические методы; модификации методов классификации; модификации методов кластеризации. Используем методы, относящиеся к двум последним группам методов.

*Обучение с подкреплением (reinforcement learning)* предполагает, что рассматриваемая стратегия (агент) обучается на основе взаимодействия с некоторой средой, которая выполняет функции учителя. Помимо стратегии и среды, обучение с подкреплением в общем случае использует еще функцию поощрения, функцию ценности и, возможно, модель среды [10]. Большое число методов решения задачи обучения с подкреплением рассмотрено, например, в работе [10]. Применяем метод, который предполагает поиск оптимальной стратегии непосредственно в пространстве стратегий без использования функции ценности.

*Активное обучение (active learning)* имеет ту особенность, что стратегия может самостоятельно назначать следующий объект для получения верного отклика, то есть самостоятельно формировать запросы к учителю (или к его модели) [11]. Полагаем, что алгоритм, реализующий активное обучение, может на шаге дообучения  $t$  сгенерировать неразмеченную выборку  $\hat{U}^t$  и предъявить ее учителю, который размечает эту выборку и возвращает результат разметки стратегии. Основной проблемой активного обучения является формирование в некотором смысле наилучшей выборки  $\hat{U}^t$ . Известно несколько подходов к определению этой выборки. Используем два из таких подходов: генерация объектов выборки  $\hat{U}^t$  в тех подобластях индикаторного пространства, плотность покрытия которых текущей выборкой  $U(t)$  минимальна; генерация

объектов выборки  $\hat{U}^t$  в окрестности тех точек текущего множества  $\hat{U}^t$ , в которых ошибка аппроксимации функции  $e(\Phi)$  имеет максимальное значение.

*Многоуровневое обучение.* Для оценки каждой из рассматриваемых метакомпетенций может быть использовано многоуровневое, точнее говоря, двухуровневое обучение. Аргументом в пользу такого подхода является высокая суммарная размерность векторов индикаторов каждой из указанных метакомпетенций. Идея подхода предполагает следующие основные этапы: 1) объединение индикаторов каждой из метакомпетенций в группы; 2) независимое решение задач обучения для каждой из указанных групп (первый уровень обучения); 3) на основе полученных результатов решение задачи обучения для интегрального критерия соответствующей метакомпетентности (второй уровень обучения).

Данный подход может быть скомбинирован со всеми рассмотренными методами обучения. Вообще говоря, для решения задач первого и второго уровней могут быть использованы различные методы. Более того, для решения различных задач первого уровня применимы не обязательно одинаковые методы.

*Многозадачное обучение (multi-task learning)* означает одновременное обучение стратегий для группы взаимосвязанных задач [12]. Для задачи определения метакомпетенций учащихся данный тип обучения может быть использован с целью решения рассмотренных выше задач обучения первого уровня, а также с целью одновременной оценки метапредметных, метакогнитивных и метакреативных компетенций. Многозадачное обучение применимо ко всем рассмотренным методам обучения.

## 6. Программная система *META*<sup>3</sup>

Разрабатываемые модели, методы и алгоритмы, обеспечивающие оценку метакомпетенций учащихся, реализуем в рамках программной системы *META*<sup>3</sup>. Выделяем два следующих режима эксплуатации системы: режим однократного обучения стратегий; режим постоянного либо эпизодического дообучения. Система разрабатывается как комплекс веб сервисов. Многокомпонентная архитектура на основе веб-сервисов позволит в дальнейшем использовать возможности облаков для гибкого масштабирования нагрузки. Основные функции системы:

- получение данных из различных источников и их предобработка;
- хранение данных и разграничение прав доступа к ним;
- обработка данных с использованием указанных методов оценки метакомпетенций учащегося.

Предусматривается извлечение исходных данных из следующих источников:

- наиболее известные и используемые социальные сети *Twitter, Facebook, Vk, Odnoklassniki, LinkedIn*;

- различные документы учащегося (курсовые работы, проекты, отчеты, статьи);
- другие системы (системы управления ВУЗом, системы тестирования, лабораторные стенды и программные системы, системы обучения).

С целью организации хранения данных используется открытая СУБД *MySQL*. Для реализации серверной части системы используем язык программирования *PHP*, для реализации интерфейсной части - *HTML*, *CSS*, *JavaScript*. Извлечение данных из социальных сетей реализуем с помощью предоставляемых этими сетями *API*.

В настоящее время написано ядро системы, управляющее содержимым *WEB*-приложения, а также реализована следующая функциональность: добавление, удаление и аутентификация пользователей; управление и разграничение правами доступа; управление группами пользователей; хранение и редактирование данных о пользователях; извлечение данных из социальных сетей *Twitter*, *Facebook*, *Vk*, *Odnoklassniki*, *Linkedin*; некоторые функции обработки этих данных; связь с онтологическими базами данных.

### Заключение

В развитие работы авторы планируют широкую апробацию разработанного и разрабатываемого модельного, методического и программного обеспечений в процессе решения реальных задач оценки метакомпетенций учащихся на основе анализа их поведения в социальных сетях.

В работе рассмотрена стационарная постановка задачи количественной оценки метакомпетенций учащихся. Однако эти оценки в процессе обучения учащегося меняются, и закономерности их изменения во времени несут важную информацию о метапотенциях учащегося. Поэтому в развитии работы авторы планируют также рассмотрение динамической постановки задачи количественной оценки метакомпетенций учащихся.

Работа выполнена при поддержке Минобрнауки России (проект 2014-14-579-0144-043).

### ЛИТЕРАТУРА

1. Пурышева Н.С., Ромашкина Н.В., Крысанова О.А. О метапредметности, методологии и других универсалиях // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2012. №1. – С. 11-17.
2. Гаврилина Е.А., Захаров М.А. Карпенко А.П. Количественная оценка метакомпетенций учащихся на основе методов машинного обучения // Наука и образование: электронное научно-техническое издание. – URL: <http://technomag.bmstu.ru/doc/764221.html>
3. SOCIOHUB.RU. – URL: <http://www.sociohub.ru/>

4. Якушев А.В., Дейкстра Л.Й., Митягин С.А. Распределенный краулер для социальных сетей на основе модели Map/Reduce // Информационно-измерительные и управляющие системы. – 2012. №11. – С. 47-53.
5. Ландэ Д.В., Снарский А.А., Пуятин В.Г. Построение терминологической сети предметной области // Регистрация, хранение и обработка данных. – 2014. № 2. С. 114-121.
6. Карпенко А.П. Меры важности концептов в семантической сети онтологической базы знаний // Наука и образование: электронное научно-техническое издание. – URL: <http://technomag.edu.ru/doc/151142.html>
7. Галямова Е.В., Карпенко А.П., Соколов Н.К, Ягудаев Г.Г. Контроль понятийных знаний субъекта обучения в обучающей системе // Вестник МАДИ (ГТУ). – 2009. №2(17). – С. 82-86.
8. Scikit-learn. Machine Learning in Python. URL: <http://scikit-learn.org/stable/>
9. Zhu Xhu. Semi-Supervised Learning Literature Survey // Computer Sciences TR 1530. – Madison: University of Wisconsin, 2008. – 60 p.
10. Саттон Р. С., Барто Э. Г. Обучение с подкреплением; пер. с англ. М.: БИНОМ. Лаборатория знаний, 2012. – 399 с.
11. Margie M. Clickers in the Classroom: An Active Learning Approach // EDUCAUSE Quarterly. 2007. – No. 2. – P. 71 - 74.
12. Varnek A. et al. Inductive Transfer of Knowledge: Application of Multi-Task Learning and Feature Net Approaches to Model Tissue-Air Partition Coefficients // Journal of Chemical Information and Modeling. 2009. Vol. 49. № 1. P. 133-144.

# **МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МНОГОМЕРНЫХ КВАЗИСТАЦИОНАРНЫХ ЭЛЕКТРОМАГНИТНЫХ ПОЛЕЙ В КАНАЛЕ ЭЛЕКТРОДИНАМИЧЕСКОГО УСКОРИТЕЛЯ**

**М.П. Галанин**

*ИПМ им. М.В. Келдыша РАН*

Необходимость в решении задач моделирования электродинамических ускорителей возникает из потребностей науки и техники. В частности, такие устройства позволяют получать уникальные скорости макротел, превышающие скорости, даваемые обычными пороховыми ускорителями, что дает возможность создавать новые приборы и устройства для исследования поведения вещества при сверхвысоких скоростях, давлениях и т.п. Эффективная разработка таких устройств, а также исследование протекающих в них явлений, без математического моделирования невозможны.

В докладе представлены методы математического моделирования квазистационарных электромагнитных полей в неоднородных областях канала ускорителя (в том числе с изменяющимися во времени, несвязными и негладкими границами подобластей), построенные и программно реализованные вычислительные алгоритмы для моделирования процесса электромагнитного ускорения в указанных областях, методами вычислительного эксперимента проведено исследование эрозии металлического контакта, а также качественных особенностей распределений электромагнитных полей в канале ускорителя в процессе разгона.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 15-01-03073).

# ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ МОДЕЛЕЙ УПРУГОПЛАСТИЧЕСКОГО ТЕЛА ДЛЯ ОПИСАНИЯ ВЫСОКОСКОРОСТНОГО ДЕФОРМИРОВАНИЯ ЛЕНТЫ ИЗ АЛЮМИНИЯ

М.П. Галанин, А.П. Лотоцкий, А.С. Родин

*ИПМ им. М.В. Келдыша РАН*

В предыдущих работах авторов [1-3] достаточно подробно рассмотрены математические и численные модели, применяемые для описания функционирования магнитного компрессора (МК), в котором магнитное поле сжимается плоским алюминиевым лайнером, ускоренным электродинамическими силами до скорости порядка 0.8 км/с.

Движение ленты лайнера исследуется в различных двумерных приближениях: поперечная модель соответствует поперечному сечению МК (плоскостью  $z=\text{const}$ ), а продольная модель – продольному сечению МК (плоскостью  $y=\text{const}$ ).

В настоящее время поставлена задача экспериментального и численного исследования поведения материала лайнера при его соударении с наклонной призмой при скорости до 1.3 км/с. Для проведения экспериментов в ГНЦ РФ ТРИНИТИ создана установка, в состав которой входит виток ускорителя и алюминиевый плоский лайнер. На рис. 1 приведена характерная схема продольного сечения устройства (красным цветом выделена лента лайнера). В системе используется емкостный источник питания. Как и в рассмотренных ранее экземплярах МК в данной конструкции виток ускорителя и лайнер последовательно соединены в одну электрическую цепь, но вывода энергии магнитного поля во внешнюю цепь не производится.

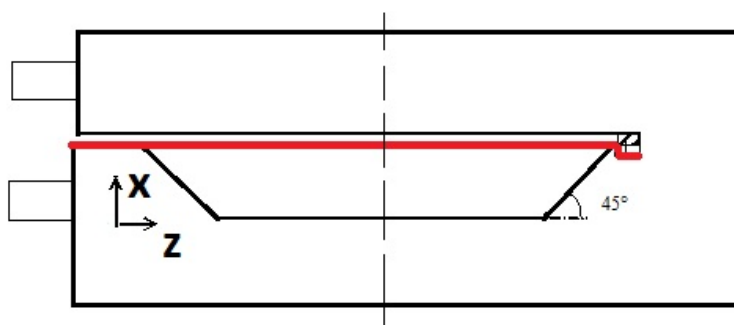


Рис. 1. Продольное сечение магнитного компрессора

Поскольку проведение трехмерных расчетов является достаточно сложной и трудоемкой задачей, то на данном этапе исследований целесообразно использовать комбинацию двумерных моделей, позволяющую

рассчитать удар ленты лайнера с кинематическими характеристиками, близкими к экспериментальным.

В качестве одного из вариантов такой комбинации предложен следующий алгоритм: поскольку в поперечной модели удается с удовлетворительной точностью рассчитать конфигурацию магнитного поля и распределение плотности тока в ленте, то по итогам соответствующего расчета можно получить график интегральной силы Лоренца, действующей на лайнер. После этого, используя данный график и предполагая, что объемная сила распределена по ленте равномерно, можно решить только кинематическую часть продольной модели (не решая электромагнитную часть, поскольку она дает несовпадающие результаты).

Из экспериментов известно, что в лайнере возникают большие деформации, поэтому ранее использовалась соответствующая модель, позволяющая описать произвольные деформации. В то же время представляет большой интерес сравнение различных моделей упругопластического тела, чтобы доказать обоснованность применения более сложной и соответственно более затратной по времени модели для решения подобных задач. В рамках данного исследования применены 3 различные механические модели упругопластического тела.

Первая модель – это стандартная модель упругопластического тела [4], которая обоснована для случая малых деформаций и малых поворотов тела.

В качестве тензора деформации применяется линейный тензор деформации Коши, компоненты которого заданы равенствами:

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial a_j} + \frac{\partial u_j}{\partial a_i} \right). \quad (1)$$

Используется теория пластического течения, которая основана на следующих предположениях:

- производную по времени тензора деформаций  $\dot{\varepsilon}$  можно представить в виде суммы упругой  $\dot{\varepsilon}^e$  и пластической  $\dot{\varepsilon}^p$  составляющих:

$$\dot{\varepsilon} = \dot{\varepsilon}^e + \dot{\varepsilon}^p; \quad (2)$$

- определяющее соотношение (закон Гука) для производной по времени от тензора напряжений Коши  $\dot{\sigma}$  имеет вид:

$$\dot{\sigma} = C^E : \dot{\varepsilon}^e = C^E : (\dot{\varepsilon} - \dot{\varepsilon}^p); \quad (3)$$

- производная тензора пластических деформаций  $\dot{\varepsilon}^p$  определяется по ассоциативному закону пластического течения:

$$\dot{\varepsilon}^p = \lambda \sigma'. \quad (4)$$

Здесь  $C^E$  - тензор коэффициентов упругости,  $\sigma'$  - девиатор тензора напряжений параметр  $\lambda$  имеет смысл множителя Лагранжа.

Во второй модели применяются тензор деформаций Грина  $E$  и второй тензор напряжений Пиолы-Кирхгоффа  $S$ , компоненты которых заданы следующими равенствами [4]:



$$E_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial a_j} + \frac{\partial u_j}{\partial a_i} + \frac{\partial u_k}{\partial a_i} \frac{\partial u_k}{\partial a_j} \right),$$

$$\sigma_{ij} = \frac{1}{J} \frac{\partial x_i}{\partial a_k} S_{kl} \frac{\partial x_j}{\partial a_l}, \quad (5)$$

где  $J = \left| \frac{\partial x_i}{\partial a_j} \right|$  - якобиан соответствующего преобразования координат.

Формальное обобщение (самый простой и напрашивающийся шаг) соотношений теории пластического течения (2)-(4) на случай больших деформаций получается заменой производных  $\dot{\varepsilon}$  и  $\dot{\sigma}$  на производные  $\dot{\mathbf{E}}$  и  $\dot{\mathbf{S}}$  (тензоры  $\varepsilon$  и  $\sigma$  заменяются на  $\mathbf{E}$  и  $\mathbf{S}$ ). В [4] отмечается, что данная модель обоснована для больших перемещений и поворотов тела, но не для больших деформаций.

В третьей модели используются не начальные, а текущие лагранжевы координаты и закон Гука для гипоупругого материала [4]:

$$\sigma^{Tr} = \mathbf{C}^E : \mathbf{d}, \quad (6)$$

в котором в качестве производной тензора напряжений Коши фигурирует производная Трусделла.

Здесь  $\mathbf{d}$  - тензор скорости деформации, компоненты которого задаются равенствами

$$d_{ij} = \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right). \quad (7)$$

Основные соотношения теории пластического течения для данной модели формулируются следующим образом [4]:

- тензор скорости деформации  $\mathbf{d}$  можно представить в виде суммы упругой  $\mathbf{d}^e$  и пластической  $\mathbf{d}^p$  составляющих:

$$\mathbf{d} = \mathbf{d}^e + \mathbf{d}^p; \quad (8)$$

- определяющее соотношение задано формулой:

$$\sigma^{Tr} = \mathbf{C}^E : \mathbf{d}^e = \mathbf{C}^E : (\mathbf{d} - \mathbf{d}^p); \quad (9)$$

- пластическая составляющая тензора скорости деформации определяется по ассоциативному закону пластического течения:

$$\mathbf{d}^p = \lambda \sigma', \quad (10)$$

Данная модель позволяет описывать произвольные деформации.

Ниже приведем результаты, полученные при применении описанных моделей. В качестве расчетной области рассмотрена половина продольного сечения (рис. 1), т.е. моделируется движение половины ленты лайнера и ее столкновение с одной опорой, которая считается абсолютно твердым телом.

На рис. 2 показана деформированная сетка в лайнере в момент времени 30 мкс, при использовании первой модели. На рисунке видно явно нефизическое утолщение ленты в области контакта с призмой. Из полученных результатов можно сделать вывод, что данная модель не применима для расчета подобных задач.

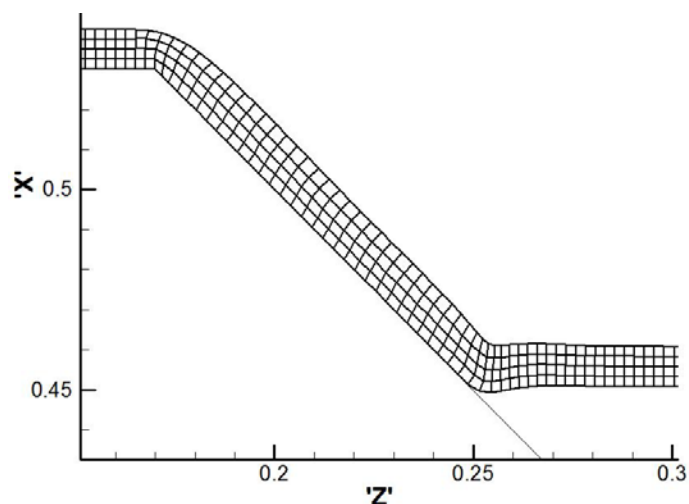


Рис. 2. Модель 1: деформированная сетка в лайнере (30 мкс)

На рис. 3 показана деформированная сетка в лайнере в момент времени 25 мкс, при использовании второй модели. Из рисунка видно, что происходит разрушение расчетной сетки (перехлест ячеек). Значит данная модель также не позволяет адекватно описывать большие пластические деформации в подобных задачах.

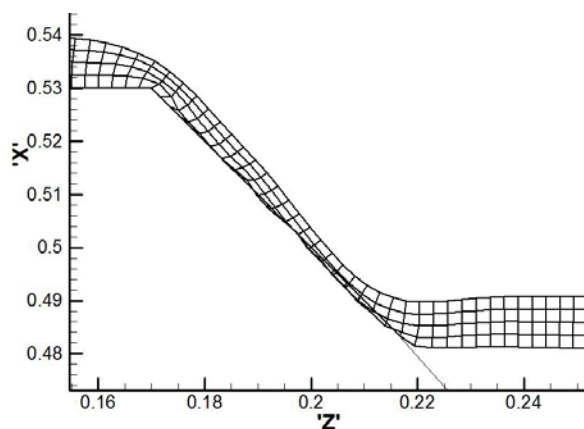


Рис. 3. Модель 2: деформированная сетка в лайнере (25 мкс)

На рис. 4 и рис. 5 показаны деформированная сетка и распределение вертикальной скорости в лайнере (нормированной на величину 100 м/с) в момент времени 40 мкс.

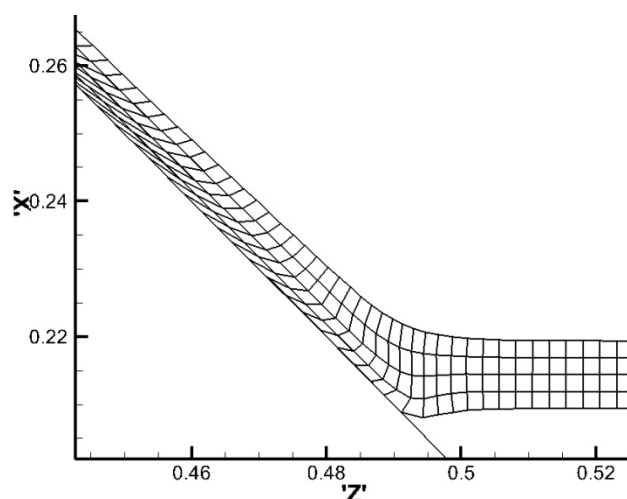


Рис. 4. Модель 3: деформированная сетка в лайнере (40 мкс)

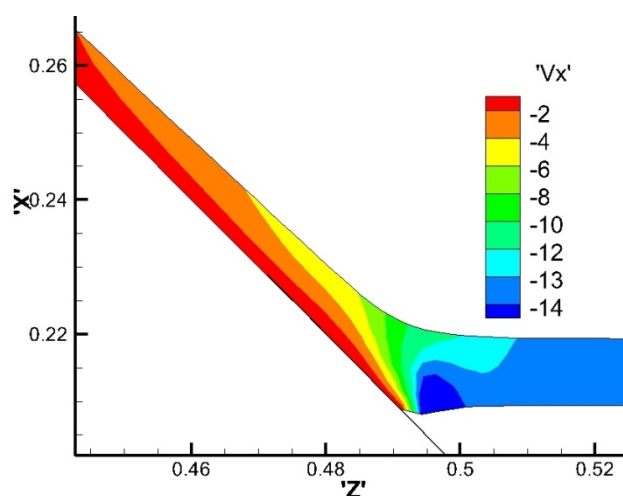


Рис. 5. Модель 3: распределение скорости в лайнере (40 мкс)

Эти результаты, полученные при применении 3 модели, достаточно хорошо согласуются с экспериментальными данными: часть ленты лайнера в процессе движения контактирует с призмой, а центральная часть ленты совершает практически плоскопараллельное движение.

Проведенные исследования позволяют сделать вывод, что в задачах высокоскоростного деформирования металлической ленты применение стандартных моделей, обоснованных для случая малых деформаций, может привести к появлению нефизических эффектов. В подобных задачах требуются более сложные модели, учитывающие большие пластические деформации.

Работа выполнена при частичной финансовой поддержке РФФИ (проекты № 14-01-31496, № 15-01-03073).

## ЛИТЕРАТУРА

1. М.П. Галанин, А.П. Лотоцкий, А.С. Родин, И.А. Щеглов. Движение лайнера в поперечном сечении магнитного компрессора // Вестник МГТУ им. Н.Э. Баумана. 2010. № 2. С. 65-84.
2. М.П. Галанин, А.П. Лотоцкий, А.С. Родин. Математическое моделирование движения лайнера в различных сечениях магнитного компрессора // Математическое моделирование. 2010. Т 22, № 10. С. 35-55.
3. Исследования работы импульсного магнитного компрессора с электродинамическим разгоном лайнера. Е.В. Грабовский [и др.] // Журнал технической физики. 2014. Т. 84. № 7. С. 126-135.
4. С.Н. Коробейников. Нелинейное деформирование твердых тел. Новосибирск: Издательство СО РАН. 2000. 262 с.

# ИНТЕГРИРОВАННАЯ ПРОГРАММНАЯ ПЛАТФОРМА СОПРОВОЖДЕНИЯ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА ТЕМЕТОС: АРХИТЕКТУРА И ЗАДАЧИ

М.П. Галанин, В.В. Лукин, А.С. Родин

*Институт прикладной математики им. М.В. Келдыша РАН*

## 1. ВВЕДЕНИЕ

Для решения широкого класса инженерных и прикладных научных задач становятся все более востребованными физически, математически и вычислительно сложные модели, требующие применения высокоточных методов [1]. Подобные методы ведут к большому объёму вычислительной работы, поэтому математическое моделирование в научно-технических приложениях требует высокой эффективности разрабатываемых алгоритмов, включая распараллеливание вычислений.

С другой стороны, многие этапы построения, программной реализации и исследования математической модели (задание геометрии области, построение сеток, ввод начальных и граничных данных модели, запуск и отслеживание работы параллельной версии расчетных процедур, визуализация и анализ полученного решения) являются общими для задач из различных областей теории и практики. Важным в такой ситуации становится создание интегрированной программной платформы, позволяющей задействовать высокоэффективные (в том числе параллельные) численные алгоритмы для решения задач различной размерности, содержащих уравнения разных типов и разного физического содержания (в рамках связанных задач).

Существует большое количество вычислительных программных платформ, позволяющих решать стандартные задачи математической физики в различных областях инженерной и прикладной научной практики. Большая часть из них является коммерческой (такие как ANSYS, Nastran, Abaqus, Fluent и др.), имеет продвинутый пользовательский интерфейс, включает пре- и постпроцессор, а также множество закрытых проблемно-ориентированных *модулей-решателей* для моделирования термомеханических, газовых и других процессов с применением метода конечных элементов (МКЭ) и (реже) метода конечных разностей (МКР). Универсальность подобных программных платформ приводит к большим трудностям при попытках решения в рамках таких пакетов физически и математически сложных задач в областях, включающих разного рода неоднородности физических параметров или геометрических характеристик.

В последнее время активно развиваются открытые программные платформы для проведения инженерных и, реже, научных расчетов (например, OpenFOAM, Salome, Code-Aster, Code Athena), создаваемые сообществами

программистов из разных стран, зачастую не объединенных в коммерческие корпорации. Подобные программные пакеты изначально строятся на открытых интерфейсах обработки данных, обеспечивая универсальность применения пакета и позволяя группам разработчиков расширять возможности платформы за счет включения собственных МКЭ или МКР расчетных кодов. В то же время в рамках подобных платформ реализованы, как правило, наиболее стандартные подходы к решению задач математической физики без учета особенностей конкретных постановок практически значимых задач. Среди таких особенностей – необходимость применения методов повышенного порядка аппроксимации, контроль консервативности расчетной схемы и др., т.е. все то, что чаще всего отличает авторские прикладные программные комплексы, разрабатываемые под специальные задачи.

В данной статье представлены результаты разработки интегрированной программной платформы для сопровождения вычислительного эксперимента в комплексных задачах математического моделирования – кроссплатформенной графической инструментальной среды Теметос. Платформа предназначена для построения и исследования математических моделей объектов и процессов, включая моделирование электромагнитных, тепловых, упругих и газодинамических полей в областях, содержащих геометрически сложные подобласти с резко неоднородными свойствами. Основная цель разработки платформы заключается в создании программного окружения, позволяющего с помощью ряда открытых интерфейсов передачи данных и управления расчетом оснастить авторский проблемно-ориентированный модуль-решатель инструментами подготовки, проведения и анализа результатов вычислительного эксперимента. В качестве базовых решателей для платформы адаптированы следующие два модуля.

1. Параллельный программный комплекс для решения двумерных задач идеальной магнитной гидродинамики разрывным методом Галеркина (RKDG-метод) [2]. Модуль используется для моделирования процессов развития магнито-ротационной неустойчивости в околосредней плазме.

2. Термомеханический решатель [3], позволяющий, в частности, исследовать прочность конструкций в рамках связанной термоупругопластической задачи с учетом разрушения и контакта. Модуль используется для численного исследования работы магнитного компрессора, а также моделирования элементов конструкций, испытывающих высокие тепловые нагрузки.

Работа выполнена при частичной финансовой поддержке РФФИ (проекты №№ 14-01-31496, 15-01-03073), а также гранта по поддержки ведущих научных школ НШ-1434.2012.2.

## 2. ОБЩАЯ АРХИТЕКТУРА ПЛАТФОРМЫ ТЕМЕТОС

Блок-схема платформы сопровождения вычислительного эксперимента состоит из групп процедур, представленных на рис. 1.

В соответствии с указанным разбиением основных процедур моделирования на группы программная платформа Теметос состоит из трех основных подсистем (см. рис. 2).

1. *Математическая подсистема.* Включает в себя основные средства проведения моделирующих расчетов, прежде всего набор программ-решателей, реализующих выбранные методы решения систем уравнений той или иной предметной области: уравнений магнитной гидродинамики, связанной термомеханики и др. Также подсистема содержит набор программ-построителей сеток (двумерных треугольных и четырехугольных; трехмерных). Подсистема состоит из набора консольных утилит. При этом утилиты могут быть как встроенными в платформу, так и исходно внешними (независимыми). Для этого разработаны интерфейсы, обеспечивающие трансформацию данных, которыми оперирует платформа, в форматы входных данных модулей-решателей.

2. *Инструментальная графическая подсистема.* Обеспечивает визуальную (трехмерную) подготовку геометрической модели рассматриваемого объекта или процесса, в случае необходимости – компоновку сложной расчетной области из простых подобластей, задание начальных и граничных условий, запуск расчета, визуализацию полученных результатов, а также взаимодействие с полным набором используемых для моделирования предметных библиотек. С программной точки зрения подсистема является кроссплатформенной, написана на языке C++ с использованием программной библиотеки оконного интерфейса Qt и технологии трехмерной визуализации VTK. Перечисленные библиотеки и технологии являются открытыми и свободнораспространяемыми.

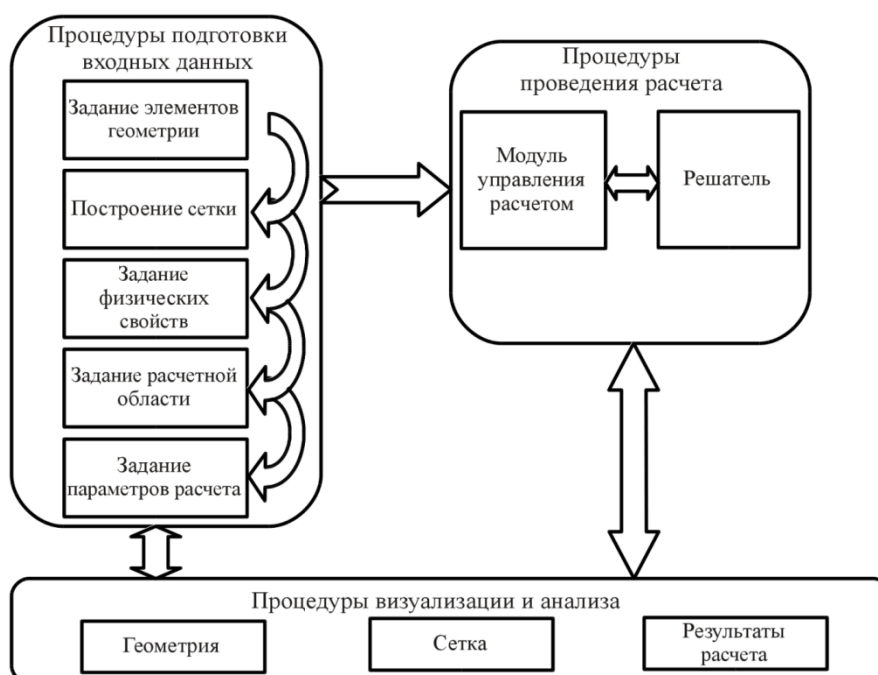


Рис. 1. Процедурная блок-схема платформы Теметос.

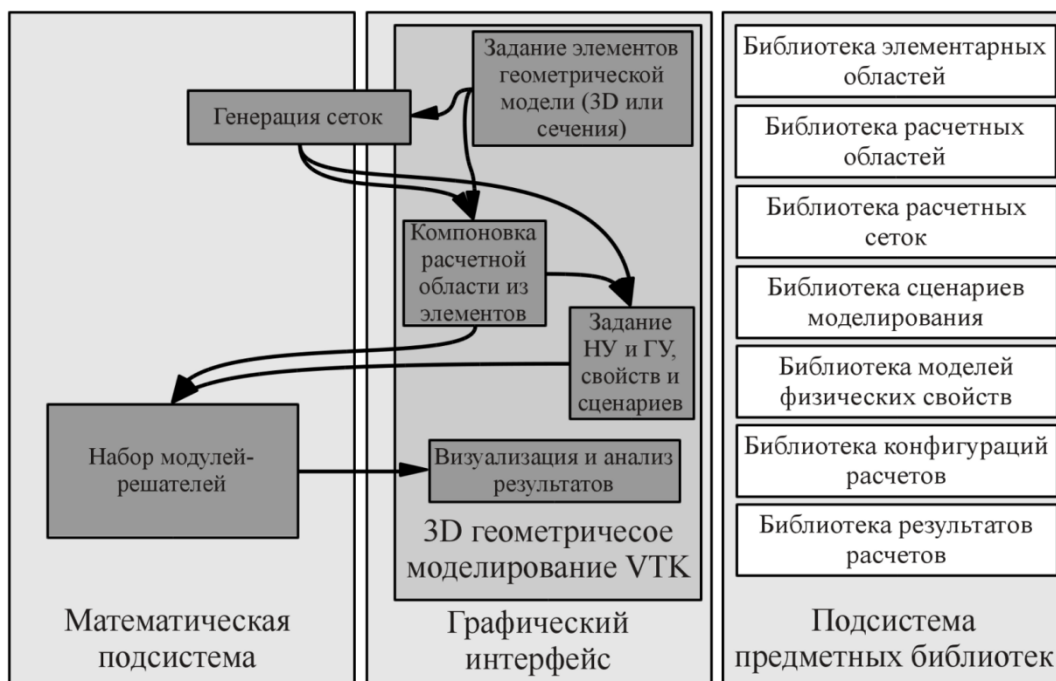


Рис. 2. Подсистемы платформы Теметос.

3. *Подсистема предметных библиотек.* Содержит структурированную пополняемую иерархическую базу знаний о моделируемых объектах и физических процессах, в частности, наборы исследуемых расчетных областей и соответствующих им сеток, реализованные в виде программных модулей модели физических свойств материалов и сред, а также средства и стандарты взаимодействия библиотек и работы с ними.

В процессе расчета взаимодействие пользователя с каждой из подсистем может осуществляться как через графическую среду, так и непосредственно. Осуществление вычислительного эксперимента для конкретной математической модели в платформе Теметос представляется в виде последовательной работы с рядом библиотек, осуществляемой в рамках единого интерфейса.

### 3. ФИЗИЧЕСКАЯ ПОДСИСТЕМА ПРЕДМЕТНЫХ БИБЛИОТЕК

Выбор программных средств при разработке интегрированной платформы сопровождения вычислительного эксперимента направлен на обеспечение возможности простого расширения набора моделируемых объектов, физических и математических моделей, используемых сеток, модулей-решателей. Для этого используется концепция безболезненного роста программы [4], причем точки роста платформы предполагается оформлять в виде предметных библиотек, которые могут содержать как текстовые файлы описания данных, так и скомпилированные программные библиотеки, подключаемые к платформе при помощи стандартизуемых интерфейсов.



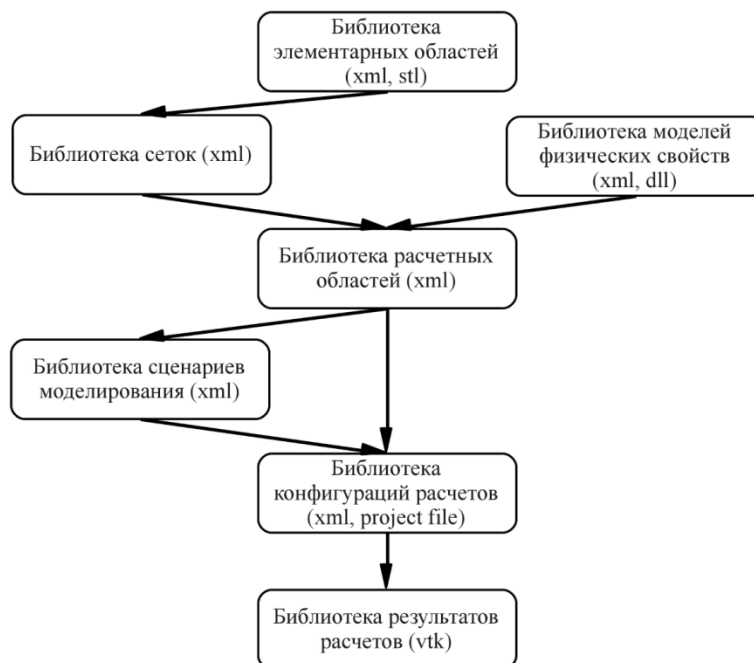


Рис. 3. Иерархия предметных библиотек.

Информационная структура платформы Теметос основана на системе расширяемых предметных библиотек, призванных охватывать базу знаний о классе моделируемых объектов и явлений. В рамках данной базы осуществляется поддержка задания геометрически элементарных составляющих области моделирования, компоновки сложных расчетных областей, реализации моделей физических свойств сред и материалов, определения сценариев моделирования (начальных и граничных условий, временных и пространственных распределений источников и стоков и проч.). Пополнение предметных библиотек позволяет описывать и моделировать полный набор исследуемых объектов. Наполнение библиотек определяется областью моделирования и, соответственно, используемым решателем.

Элементы библиотек связаны друг с другом принципом соответствия, подобным применяемому в реляционных базах данных. Разработана система работы с набором предметных библиотек, позволяющая на каждой стадии подготовки расчета конструкции получать полную информацию о всех составляющих элементах модели.

В состав платформы входят следующие библиотеки.

1. Геометрическая библиотека элементарных областей (БЭО), содержащая наиболее общее (векторное) описание используемых при моделировании геометрических областей (в случае инженерных расчетов – элементов конструкций: лопаток, шестерней и проч.).
2. Библиотека сеток (БС), содержащая сетки, заданные в элементарных областях из БЭО.
3. Библиотека моделей физических свойств (БМФС), включающая как текстовое описание физических свойств моделируемых материалов и

сред в простых случаях, так и скомпилированные dll-библиотеки подпрограмм, описывающих сложное поведение параметров среды (теплопроводности, пористости, электрической проводимости и др.); набор доступных для описания физических параметров определяется возможностями конкретного модуля-решателя.

4. Библиотека расчетных областей (БРО), содержащая скомпонованные расчетные области, состоящие из относительно простых элементов, хранящихся в БЭО, вместе с выбранными для каждого элемента сеткой из БС и заданными моделями среды из БМФС.
5. Библиотека сценариев моделирования (БСМ), содержащая наборы входных данных, начальных и граничных условий, временных и пространственных распределений источников для расчета объекта, компоновка которого содержится в БРО.
6. Библиотека конфигураций расчетов (БКР), содержащая наборы настроек модуля-решателя, включая составляющие решаемых систем уравнений, настройки численных методов и алгоритмов (в том числе параллельных), применяемых к расчету объекта из БРО с заданным сценарием моделирования из БСМ.
7. Библиотека результатов расчетов (БРР), содержащая полный набор выходных данных решателя при проведении расчета из БКР.

Иерархические связи предметных библиотек показаны на рис. 3.

#### 4. ИНСТРУМЕНТАЛЬНАЯ ГРАФИЧЕСКАЯ ПОДСИСТЕМА И АЛГОРИТМ РАБОТЫ С ПЛАТФОРМОЙ

Задача платформы Теметос – оснастить исследователя, разрабатывающего собственный или использующего готовый проблемно-ориентированный модуль-решатель, максимально удобными (прежде всего, графическими) средствами подготовки и проведения вычислительного эксперимента с применением этого решателя. В состав платформы включены графические редакторы предметных библиотек, а также инструменты, позволяющие «собрать» из заранее подготовленных единиц (геометрических моделей, моделей физических свойств, сеток, конфигураций решателя) расчет, произвести его запуск и отслеживание, проанализировать его результаты.

Описание геометрии расчетной области исследуемого объекта ведется на различных, но взаимоувязанных, уровнях. Полная расчетная область описывается в два этапа:

1. задание чертежей элементарных областей в виде трехмерных или плоских областей (с учетом осевой симметрии) с указанием всех размеров и подробным заданием формы каждого элемента (рис. 4);

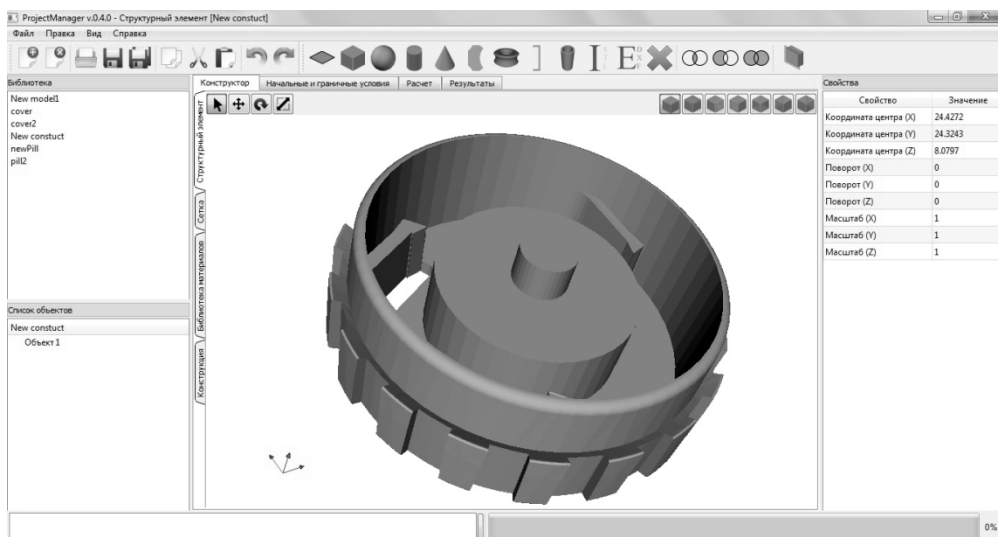


Рис. 4. Редактор элементарных трехмерных областей платформы Теметос.

2. формирование из предварительно подготовленного набора элементарных областей своеобразной схемы-сборки расчетной области с указанием положения и набора физических моделей для каждого из элементов.

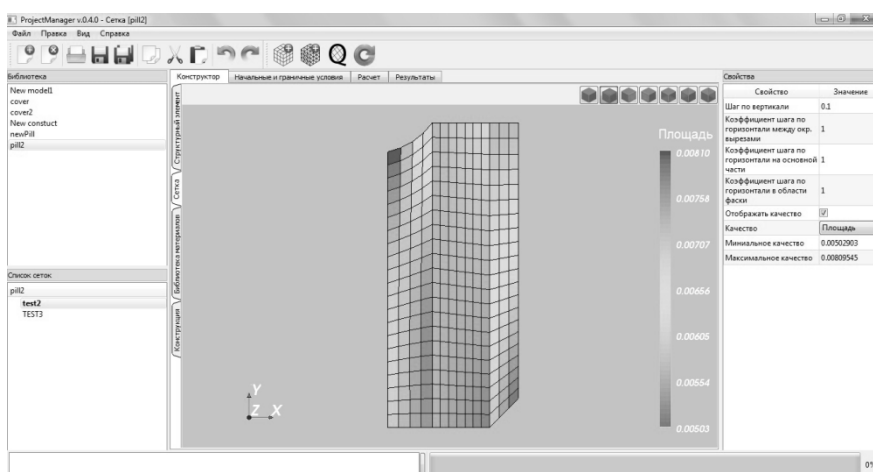


Рис. 5. Задание треугольной сетки в произвольной области.

Подобный подход оказывается удобным в ряде случаев моделирования конструкций. В частности, численный алгоритм термомеханического решателя с учетом контакта и мультиконтакта упругих тел позволяет моделировать различные элементы инженерной конструкции как независимые твердые тела, взаимодействующие через поверхности контакта. Это позволяет независимо задавать геометрию, физические и модельные свойства каждого элемента, и моделирование конструкции в целом начинается именно с проектирования (введения в соответствующие библиотеки) индивидуальных элементов с наиболее подробным учетом особенностей их формы. Реализован импорт трехмерных областей, заданных в формате STL, импорт в который осуществляется из многих популярных CAD-систем.

Для каждой элементарной подобласти пользователь может задать одну или несколько сеток (рис. 5), причем одна и та же подобласть может входить в состав расчетной области многократно, имея при этом разные сетки. Платформа Теметос позволяет визуально и численно (используя ряд критериев) оценить качество сетки и при необходимости провести ее оптимизацию.

Далее пользователь производит выбор физических моделей сред и материалов, которые будут использоваться в расчете. Каждая среда представляется набором физических свойств, каждое из которых может быть задано либо непосредственно численно, либо указанием функции в подключаемой dll-библиотеке. Модуль-решатель должен поддерживать подключение внешних библиотек. В рамках работы над платформой разработан унифицированный интерфейс подключения материальных библиотек, который может быть реализован в виде API платформы, доступного для использования во внешних решателях.

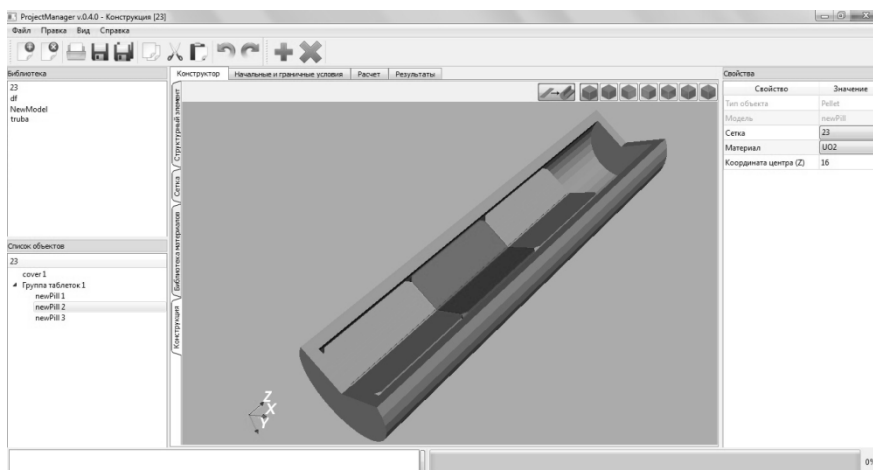


Рис. 6. Редактор расчетных областей – трехмерная форма отображения конструкции.

Из подготовленных элементарных подобластей с выбором для каждой из них сетки и набора физических свойств осуществляется сборка расчетной области в соответствующем редакторе (см. рис. 6). Далее для выбранной расчетной области (конструкции) задаются начальные и граничные условия, сценарий моделирования, распределения объемных и поверхностных источников. Для реализации этой возможности создан редактор сценариев, включающий парсер формул библиотеки VTK.

После того, как задана расчетная область, определены используемые модели материалов и сценарий моделирования, производится настройка и запуск на расчет решателя. В соответствующем блоке платформы задаются параметры численного метода, используемые классы физических и математических моделей, осуществляется выбор моделируемой конструкции, соответствующего ей сценария нагружения, а также общие параметры расчета,

такие как конечное время, размерность задачи и проч. В процессе расчета предоставляется соответствующая телеметрическая информация.

Доступно как двумерное, так и трехмерное (в т.ч. двумерное осесимметричное) отображение результатов расчетов. Кроме того разработаны средства по трансформации способа отображения результатов – инструменты по созданию контурных графиков распределений физических величин, отображению векторных полей, соответствующих переменным задачи, а также генерации результирующей формы конструкции путем смещения узлов сетки на соответствующий вектор перемещения (см. рис. 7).

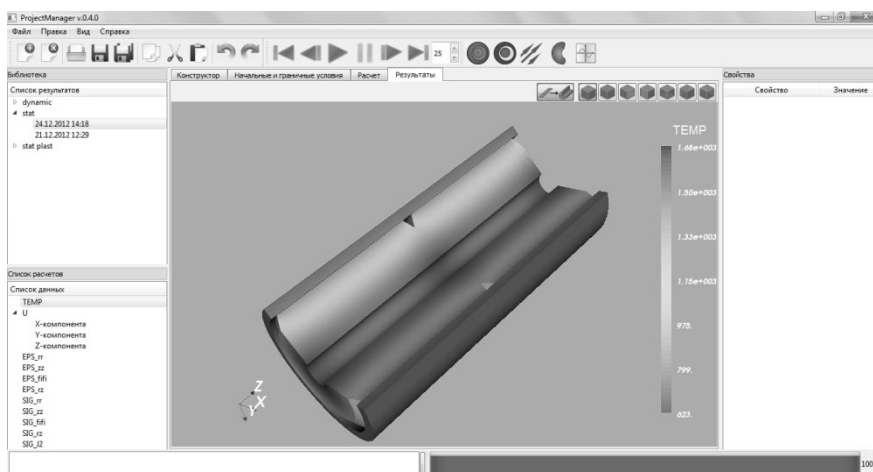


Рис. 7. Окно представления результатов расчетов.

## 5. МАТЕМАТИЧЕСКАЯ ПОДСИСТЕМА

Математическая подсистема платформы Теметос включает основные расчетные программы и комплексы, используемые для проведения вычислительного эксперимента – программы-построители сеток и модули-решатели, реализующие методы численного исследования систем уравнений, соответствующих выбранной физической задаче. Архитектурой платформы Теметос предусмотрено, что каждая расчетная программа является независимой, обмен данными между расчетной программой и платформой производится только через файлы входных данных и настроек, при наличии входных файлов расчетная программа может быть запущена непосредственно из командной строки вне графической оболочки.

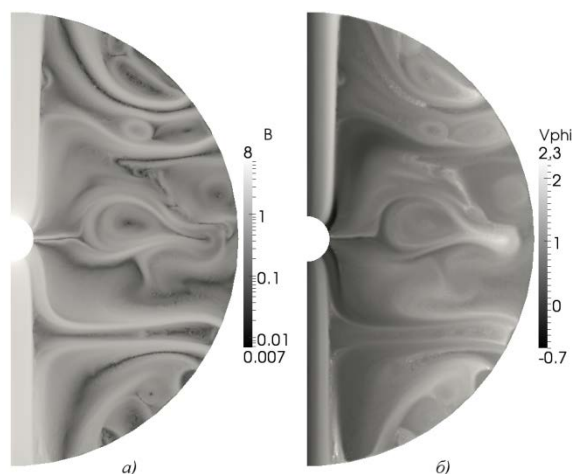


Рис. 8. Моделирование МРН в аккрецирующей плазме: модуль магнитного поля (а) и азимутальная скорость течения плазмы (б)

Платформа Теметос включает несколько строителей неструктурированных и структурированных сеток, причем каждый из них является независимым программным продуктом, а платформа содержит необходимые конвертеры для преобразования данных о геометрии в формат строителя и получаемых данных о сетке в другие используемые форматы. Поддерживается построение как неструктурированных треугольных (программа Gridder 2D [5]), так и четырехугольных сеток различных видов – как в областях общего вида, так и в специальных областях, соответствующих специфике задачи (параметрически заданные топливные таблетки в задаче о моделировании твэла).

На пилотном этапе разработке к платформе адаптируются два расчетных модуля – параллельный программный комплекс для решения двумерных задач идеальной магнитной гидродинамики RKDG-методом и термомеханический программный комплекс, включающий математические модели термоупругопластичного материала и хрупкого разрушения. Платформа применяется для задания расчетной области, запуска последовательного расчета и анализа получаемых результатов. Предполагается, что в дальнейшем в состав комплекса Теметос будут включены инструменты подготовки и проведения расчета на параллельных вычислительных комплексах, включая кластерные системы (К-100 ИПМ им. М.В. Келдыша РАН, УЭВК кафедры «Прикладная математика» МГТУ им. Н.Э. Баумана [6] и др.).

Результаты расчетов в двумерной осесимметричной постановке для задач о развитии магниторотационной неустойчивости (МРН) в аккрецирующей околозвездной плазме (постановку задачи см. в [7]).

## 6. ЗАКЛЮЧЕНИЕ

Представлена интегрированная программная платформа для сопровождения вычислительного эксперимента в комплексных задачах

математического моделирования. Предложена архитектура платформы, состоящая из трех подсистем различного назначения – математической, графическо-инструментальной и подсистемы семи предметных библиотек. Оформление в виде пополняемых библиотек элементов исследуемой модели – расчетных областей, физических свойств сред, сценариев моделирования – открывает возможности безболезненного расширения платформы. При этом каждая такая библиотека требует создания форматов данных и интерфейсов подключения внешних, не зависящих от платформы Теметос, потоков данных и процедур.

## ЛИТЕРАТУРА

1. Галанин М.П., Савенков Е.Б. Методы численного анализа математических моделей. М.: Изд-во МГТУ им. Н. Э. Баумана, 2010. 591 с.
2. Галанин М.П., Лукин В.В., Шаповалов К.Л. Параллельный алгоритм RKDG метода второго порядка для решения двумерных уравнений идеальной магнитной гидродинамики // Параллельные вычислительные технологии (ПаВТ'2013): труды международной научной конференции (1-5 апреля 2013 г., г. Челябинск). Челябинск: Издательский центр ЮУрГУ. 2013. С. 116-126.
3. Богатырь С.М., Галанин М.П., Крупкин А.В., Кузнецов В.И., Лукин В.В., Новиков В.В., Родин А.С., Станкевич И.В., Яковлев М.Е. Математическое моделирование термоупругого контактного взаимодействия осесимметричных тел. Инженерный журнал: наука и инновации, 2013, вып. 4.
4. Горбунов-Посадов М.М. Как растет программа // ИПМ им. М.В. Келдыша РАН. Режим доступа: <http://www.keldysh.ru/gorbunov/grow.htm> (дата обращения: 29.05.2013).
5. Щеглов И.А. Программа для триангуляции сложных двумерных областей Gridder2d. // Препринт ИПМ им. М.В.Келдыша РАН №60. 2008. 32 с.
6. Лукин В.В., Марчевский И.К., Морева В.С., Попов А.Ю., Шаповалов К.Л., Щеглов Г.А. Учебно-экспериментальный вычислительный кластер. Ч.2. Примеры решения задач // Вестник МГТУ им. Н.Э. Баумана. Естественные науки. 2012. № 4. С. 82-102.
7. Велихов Е. П., Сычугов К. Р., Чечеткин В. М., Луговский А. Ю., Колдоба А.В. Магниторотационная неустойчивость в аккрецирующей оболочке протозвезды и образование крупномасштабной структуры магнитного поля // Астрономический журнал. 2012. Т. 89, №. 2. С. 107-119.

# МУЛЬТИМЕДИЙНЫЕ ИЛЛЮСТРАЦИИ В НАУЧНОЙ ПУБЛИКАЦИИ

М.М. Горбунов-Посадов, А.В. Ермаков, Т.А. Полилова

*ИПМ им.М.В.Келдыша РАН*

Интернет уверенно превращается в основной источник научного знания. Научные журналы, монографии, энциклопедии либо получают полнотекстовую интернет-проекцию, на которую с удовольствием переключается подавляющее большинство читателей, либо и вовсе целиком переключиваются в интернет, постепенно забывая о своих печатных корнях.

Среди многочисленных преимуществ размещения научной публикации в интернете далеко не последнюю роль играет возможность широко использовать звук, видео, анимацию, интерактивную компьютерную графику — богатейший набор средств, о которых не могли и мечтать печатные издания. К сожалению, эти средства мультимедиа включаются в научные публикации недостаточно часто.

А ведь технический арсенал мультимедиа учеными давно освоен. Редкий семинар или защита диссертации проходят сегодня без подготовленной докладчиком динамичной демонстрации сталкивающихся галактик, движущихся циклонов, совершающих головокружительные манипуляции роботов и т. д. Почему же эти замечательные иллюстрации до сих пор не украшают собой публикуемые научные работы?

Причин несколько. Прежде всего, инерция представления о публикации как о чем-то абсолютно неподвижном, что обязано в полном объеме отразиться на бумаге. До сих пор чрезвычайно редко встречаются научные журналы, где автору статьи разрешается подготовить мультимедиа иллюстрации для онлайн-версии. Как ни странно, даже если журнал исключительно онлайн-версии, тем не менее мультимедийные иллюстрации в нем зачастую не допускаются.

Существуют и чисто технические причины, мешающие массовому появлению мультимедиа иллюстраций в научной статье. В то время как у создателя "ненаучного" сайта с появлением HTML5 просто глаза разбегаются от обилия всевозможных доступных приемов работы с мультимедиа, обустройство "научного" интернета вызывает определенные сложности. Дело в том, что для представления текстов научных статей HTML практически не используется, там безраздельно господствует формат PDF. А организация размещения мультимедиа в PDF коренным образом отличается от технологий HTML, причем, увы, не в лучшую сторону.

Анализируется опыт использования видео, анимации и 3D иллюстраций в онлайн-издании "Препринты ИПМ им.М.В.Келдыша".



Работа поддержана грантом РФФИ 13-01-00493-а.

## ЛИТЕРАТУРА

1. М.М.Горбунов-Посадов, Д.С.Ролдугин, М.И.Слепенков, И.В.Тузов. Анимация и видео в научной публикации // Препринты ИПМ им. М.В.Келдыша. 2014. № 104. 32 с. URL: <http://library.keldysh.ru/preprint.asp?id=2014-104>
2. М.М.Горбунов-Посадов. Оформление препринта ИПМ им.М.В.Келдыша. — URL: <http://keldysh.ru/preprints/sample.doc>

# СЕМАНТИЧЕСКОЕ АННОТИРОВАНИЕ В СИСТЕМЕ УПРАВЛЕНИЯ ФИЗИКО-МАТЕМАТИЧЕСКИМ КОНТЕНТОМ

А. М. Елизаров<sup>1</sup>, Н. Г. Жильцов<sup>2</sup>, А. В. Кириллович<sup>2</sup>, Е. К. Липачёв<sup>1</sup>

*1 Институт математики и механики им. Н. И. Лобачевского  
Казанского (Приволжского) федерального университета*

*2 Высшая школа информационных технологий и информационных систем  
Казанского (Приволжского) федерального университета*

**Введение.** В настоящее время для научного сообщества одной из важных проблем, связанных с информатизацией исследовательской деятельности, является организация индивидуального информационного цифрового пространства ученого [1], а наиболее эффективный подход к решению проблем управления научной информацией обеспечивают семантические технологии, причем в результате автоматизации процессов обработки семантических связей формируется личное информационное пространство ученого.

Развитие методов автоматической обработки текстов позволило решать задачи извлечения знаний в терминах онтологий [2]. Онтологические модели предметных областей используются как технологическая основа построения рекомендательного сервиса, позволяющего выполнить персонализированный отбор научных документов в соответствии с семантическим профилем учёного [3].

**Семантическое аннотирование.** В настоящее время на платформе Science.tatarstan.ru (<http://science.tatarstan.ru/>) создается система аннотирования, основанная на онтологиях предметных областей. В исходный текст научной работы система автоматически добавляет ссылки на определения терминов – результатом является документ, содержащий помимо исходного текста научной работы блоки аннотаций, доступные через интерфейс пользователя. Отметим в связи с этим платформу ScienceWISE (<http://sciencewise.info>), с помощью которой пользователи ежедневно получают научную информацию, упорядоченную в соответствии с их личными интересам. При этом научные статьи ранжируются в соответствии с интересами пользователей, а на основе онтологий создается система закладок и аннотаций [4].

Подход, развиваемый нами, основан на использовании онтологий естественно-научных областей, прежде всего, онтологии математических знаний OntoMath<sup>PRO</sup> [2, 5]. При формировании этой онтологии использовались методы онтологического моделирования с применением алгоритмов автоматического выделения понятий и связей [6]. Существенное значение для внедрения подобной системы имеет полнота «покрытия» онтологией терминологического массива электронной коллекции. В настоящее время в полуавтоматическом режиме проводится пополнение онтологии на основе

терминологической базы Математической энциклопедии [7 – 9] и аннотаций, представленных на портале MathNet.ru [10].

**Архитектура системы** сформирована из нескольких модулей и представлена на рис. 1.

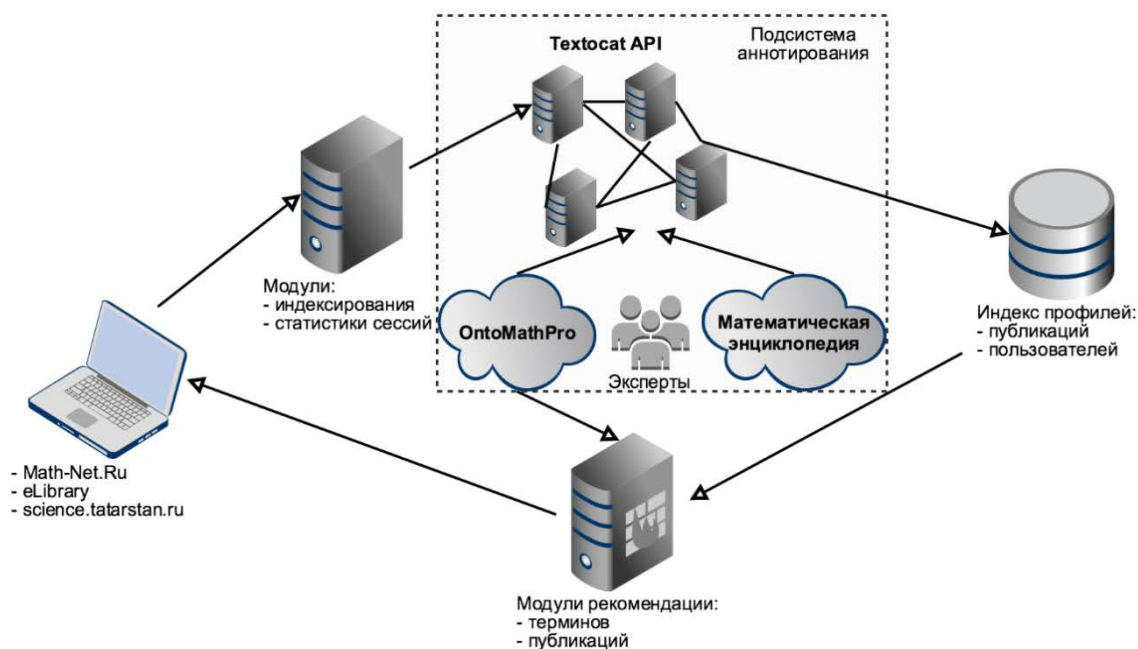


Рис. 1. Архитектура системы

Модуль семантического аннотирования производит связывание фраз из текста научных публикаций с концептами из терминологических источников, в качестве которых используются онтологии OntoMath<sup>PRO</sup> и ScienceWISE, а также база знаний DBpedia. Сначала с помощью информации, полученной на основе лексико-синтаксического анализа, реализованного в облачном сервисе текстовой аналитики Textocat (<http://textocat.com>), генерируется набор фраз, рассматриваемых как кандидаты на связывание. При проведении такого анализа используются функции точного сопоставления по словарю, а также возможности выделения границ именных групп. При этом словарь составляется из имен онтологических концептов, включая синонимы, сокращения и эквивалентные понятия.

На следующем этапе модуль ранжирует концепты – кандидаты на связывание – с полученными фразами. При ранжировании учитываются характеристики, выражающие строковую близость имен кандидатов, а также контекст слов и других выделенных концептов, окружающих фразу-кандидата.

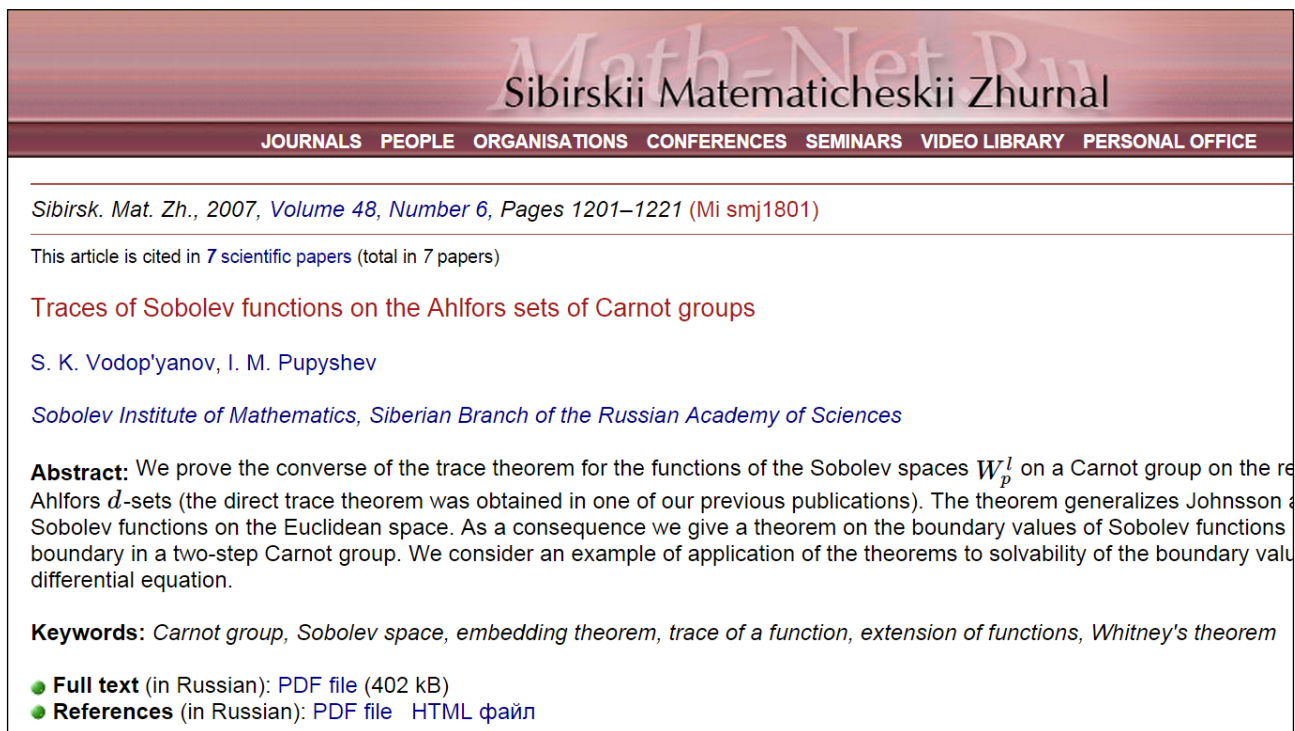
Далее происходит валидация высоко ранжированных кандидатов – каждому концепт-кандидату приписывается свое ранжирующее значение, рассматриваемое как «уверенность». В случае низкой уверенности модуля в кандидате на связывание для данной фразы связывание не производится.

Модуль составления семантического профиля пользователя формирует

семантический профиль на основе онтологических концептов, исходя из истории просмотра пользователем статей на сформированном ресурсе: профиль содержит основные концепты, с которым пользователю приходится работать наиболее часто, и отражает его компетенции.

Модуль рекомендации определений решает задачу составления рекомендаций при просмотре пользователем научных публикаций. Результатом являются список терминов, которые система, на основании сопоставления семантического профиля статьи и семантического профиля пользователя, выделила как наименее понятные пользователю.

**Пользовательский интерфейс** базируется на карточках публикаций (рис. 2), сформированных на портале MathNet.ru. Карточка публикации содержит название статьи, авторов, метаданные журнала, аннотацию статьи и, иногда, указанный автором список ключевых слов. Разработан специальный скрипт, результатом действия которого является новый раздел *Extracted keywords*, содержащий автоматически извлеченные ключевые слова со ссылками на карточки терминов (рис. 3 и 4).



The image shows a screenshot of a publication card on the MathNet.ru portal. The card has a dark red header with the text 'Math-Net.Ru' and 'Sibirskii Matematicheskii Zhurnal'. Below the header is a navigation bar with links: 'JOURNALS', 'PEOPLE', 'ORGANISATIONS', 'CONFERENCES', 'SEMINARS', 'VIDEO LIBRARY', and 'PERSONAL OFFICE'. The main content of the card includes the following text:

*Sibirsk. Mat. Zh.*, 2007, *Volume 48, Number 6, Pages 1201–1221* (Mi smj1801)

This article is cited in [7 scientific papers](#) (total in 7 papers)

**Traces of Sobolev functions on the Ahlfors sets of Carnot groups**

S. K. Vodop'yanov, I. M. Pupyshev

*Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences*

**Abstract:** We prove the converse of the trace theorem for the functions of the Sobolev spaces  $W_p^l$  on a Carnot group on the re Ahlfors  $d$ -sets (the direct trace theorem was obtained in one of our previous publications). The theorem generalizes Johnsson and Sobolev functions on the Euclidean space. As a consequence we give a theorem on the boundary values of Sobolev functions boundary in a two-step Carnot group. We consider an example of application of the theorems to solvability of the boundary value differential equation.

**Keywords:** *Carnot group, Sobolev space, embedding theorem, trace of a function, extension of functions, Whitney's theorem*

● **Full text** (in Russian): [PDF file](#) (402 kB)  
● **References** (in Russian): [PDF file](#) [HTML файл](#)

Рис. 2. Одна из карточек публикации на портале MathNet.ru

Отметим, что список ключевых слов имеет ряд ограничений. Во-первых, одно и то же понятие в разных статьях может обозначаться разными ключевыми словами. Во-вторых, ключевые слова не являются активными ссылками и не ведут на карточку соответствующего математического понятия. Наконец, во многих статьях список ключевых слов является неполным или даже вовсе отсутствует.

Math-Net.Ru  
Sibirskii Matematicheskii Zhurnal

JOURNALS PEOPLE ORGANISATIONS CONFERENCES SEMINARS VIDEO LIBRARY PERSONAL OFFICE

---

*Sibirsk. Mat. Zh.*, 2007, Volume 48, Number 6, Pages 1201–1221 (Mi smj1801)

This article is cited in [7 scientific papers](#) (total in 7 papers)

**Traces of Sobolev functions on the Ahlfors sets of Carnot groups**

S. K. Vodop'yanov, I. M. Pupyshev

*Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences*

**Abstract:** We prove the converse of the trace theorem for the functions of the Sobolev spaces  $W_p^1$  on a Carnot group on the Ahlfors  $d$ -sets (the direct trace theorem was obtained in one of our previous publications). The theorem generalizes Johnsson Sobolev functions on the Euclidean space. As a consequence we give a theorem on the boundary values of Sobolev functions boundary in a two-step Carnot group. We consider an example of application of the theorems to solvability of the boundary value differential equation.

**Extracted keywords:** [Sobolev space](#), [Euclidean space](#), [Boundary value](#), [Partial differential equation](#), [Smooth boundary](#)

**Keywords:** *Carnot group, Sobolev space, embedding theorem, trace of a function, extension of functions, Whitney's theorem*

● **Full text** (in Russian): [PDF file](#) (402 kB)  
 ● **References** (in Russian): [PDF file](#) [HTML файл](#)

Рис. 3. Результат работы сервиса – сформированный раздел Extracted keywords

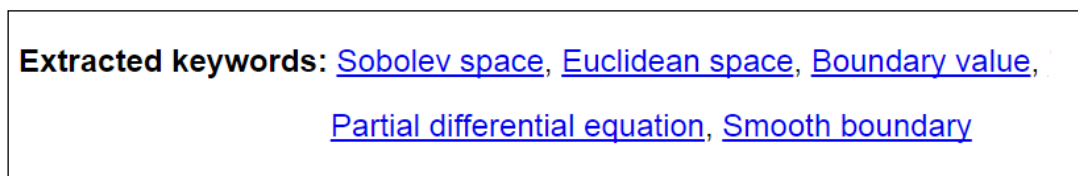


Рис. 4. Раздел Extracted keywords – автоматически извлеченные ключевые слова со ссылками на карточки терминов

Карточка термина представлена на рис. 5. Она содержит метаданные, различные наименования термина (на русском и английском языках), определение термина, ссылки на гипонимы/гиперонимы термина, ссылки на страницу термина на внешних ресурсах, а также ссылки на статьи с портала MathNet.ru, в которых использован данный термин.

Программная надстройка автоматически извлекает ключевые слова и добавляет блок автоматически извлеченных ключевых слов на карточку публикации. Каждое ключевое слово представляет собой активную ссылку на карточку соответствующего понятия. Ссылка ведет на одно и то же понятие, вне зависимости от того, какой конкретно термин использовался для обозначения этого понятия, и на каком языке.

Если авторы не указали список ключевых слов, то блок автоматически извлеченных ключевых слов является единственной заменой этого списка. Если указали, то его дополнением.

Нормированное пространство

↑

## Пространство Соболева

↓

Пространство Соболева дробного порядка, Пространство Соболева бесконечного порядка

**Определение:** Пространство функций, определенных на открытом множестве  $\Omega$  и интегрируемых с  $p$ -й степенью их модуля вместе со своими обобщенными производными до порядка  $m$  включительно.

**Внешние ресурсы:** OntoMath, ScienceWISE, MathWorld, Математическая энциклопедия, Википедия

**Публикации:**

- Л. Д. Кудрявцев, С. М. Никольский. Пространства дифференцируемых функций многих переменных и теоремы вложения // Анализ – 3, Итоги науки и техн. Сер. Современ. пробл. мат. Фундам. направления, 26, ВИНТИ, М., 1988
- А. А. Васильева. Достаточные условия вложения весового класса Соболева на области с условием Джона // Сиб. матем. журн., 56:1 (2015)
- С.К. Водопьянов, И.М. Пупышев. Следы функций из пространства Соболева на множествах Альфорса групп Карно // Сиб. матем. журн., 2007, том 48, номер 6
- Б.В. Трушин. Вложение пространства Соболева в пространство Орлича для области с нерегулярной границей // Матем. заметки, 2006, том 79, выпуск 5

Рис. 5. Карточка одного из терминов

Работа выполнена при финансовой поддержке РФФИ (проекты 15-07-08522, 15-47-02472).

## ЛИТЕРАТУРА

1. Моисеев Е.И., Муромский А.А., Тучкова Н.П., Меденников А.М. Организация персонального информационного пространства на основе модели предметной области автора // Научный сервис в сети Интернет, многообразии суперкомпьютерных миров: труды Международной суперкомпьютерной конференции (г. Новороссийск, 22-27 сентября 2014 г.). – М.: Изд-во МГУ, 2014. – С. 429-431. – URL: <http://agora.guru.ru/abrau2014/pdf/429.pdf>.
2. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G. Mathematical knowledge representation: semantic models and formalisms // Lobachevskii J. of Mathematics. – 2014. – V. 35, No 4. – P. 347-353.
3. Елизаров А.М., Жильцов Н.Г., Иванов В.В., Кириллович А.В., Липачёв Е.К., Невзорова О.А. Семантический рекомендательный сервис в профессиональной деятельности математика // Ученые записки Института социальных и гуманитарных знаний. – 2015. – Вып. 1 (13). – С. 190-197.
4. Astafiev A., Prokofyev R., Guéret C., Boyarsky A., Ruchayskiy O. ScienceWISE: A Web-based interactive semantic platform for paper annotation and ontology editing // In: Extended Semantic Web Conference 2012, Greece, 2012. URL:

[http://2012.eswc-conferences.org/sites/default/files/eswc2012\\_submission\\_316.pdf](http://2012.eswc-conferences.org/sites/default/files/eswc2012_submission_316.pdf).

5. Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E. OntoMathPRO ontology: a linked data hub for mathematics // In Knowledge Engineering and the Semantic Web. Springer International Publishing. – 2014. – P. 105-119.
6. Биряльцев Е.В., Елизаров А.М., Жильцов Н.Г., Липачёв Е.К., Невзорова О.А., Соловьев В.Д. Методы анализа семантических данных математических электронных коллекций // Научно-техническая информация. Сер. 2. Информ. процессы и системы. – 2014. – №4. – С. 12-17.
7. Математическая энциклопедия. – М.: Советская энциклопедия. – Т. 1. – 1977; Т. 2. – 1979; Т. 3. – 1982; Т. 4. – 1984; Т. 5. – 1985.
8. Математическая энциклопедия. URL: [http://gufo.me/matenc\\_a](http://gufo.me/matenc_a)
9. Encyclopedia of Mathematics. URL: [http://www.encyclopediaofmath.org/index.php/Main\\_Page](http://www.encyclopediaofmath.org/index.php/Main_Page)
10. Общероссийский математический портал. URL: <http://www.mathnet.ru/>

# РАЗРАБОТКА ОНЛАЙН-РЕПОЗИТОРИЯ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ ОПТИМИЗАЦИИ

Н.М. Ершов, Н.Н. Попова

*Московский государственный университет им. М.В. Ломоносова*

Целью проводимых исследований является разработка программной системы для автоматической генерации параллельных алгоритмов машинного обучения и эволюционных вычислений [1]. Одной из составляющих этой системы должен быть специальный веб-сервис, предназначенный для автоматической генерации параллельных программ, реализующих стандартные эволюционные алгоритмы оптимизации. В настоящей работе описывается прототип такого сервиса, ориентированный на генерацию последовательного кода классических эволюционных оптимизационных алгоритмов, таких как генетические алгоритмы [2, 3], метод роя частиц [4], муравьиные алгоритмы [5] и т.п.

Отличительной особенностью этих методов является то, что все они являются метаэвристиками: каждый из этих алгоритмов представляет собой, по сути, целое семейство алгоритмов, объединенных некоторой общей идеей, но сильно отличающихся друг от друга на нижнем уровне (например, реализацией тех или иных операторов). Целью работы разрабатываемой онлайн-службы является предоставление пользователю возможности собрать свой вариант оптимизационного алгоритма, используя для этого набор стандартных схем и элементов. Результатом работы системы должен быть программный код (например, на языке программирования C++) выбранного и настроенного пользователем алгоритма оптимизации. Этот код должен а) быть работающим; б) быть хорошо задокументированным; в) содержать минимальный набор необходимых вспомогательных структур и функций (без использования сложных классов или шаблонов классов и функций). У пользователя системы, таким образом, появляется полный доступ к полученному коду, он может использовать его, например, в качестве некоторой заготовки для разработки своей собственной реализации того или иного алгоритма.

Архитектура всей системы является относительно простой (рис. 1). Вся информация о доступных пользователю схемах и элементах хранится в формате JSON в отдельном конфигурационном файле. Эта информация используется в частности веб-страницей для организации диалога с пользователем, который реализуется в формате помощника (wizard) – на каждом шаге диалога пользователю предлагается выбрать одну из возможных опций текущего элемента (рис. 2). Собранная информация отправляется на сервер, где специальный скрипт (по сути, ядро всей системы) генерирует требуемый код, используя для этого библиотеку готовых элементов. Сгенерированный код алгоритма возвращается клиенту в простом текстовом формате.





Рис. 1. Архитектура системы

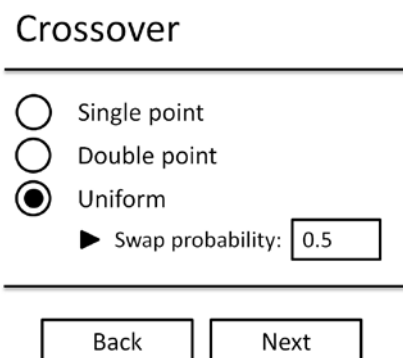


Рис. 2. Элемент интерфейса системы

Наиболее нетривиальной частью рассматриваемой системы является взаимная организация отдельных элементов эволюционных алгоритмов. Каждый алгоритм представляется набором шаблонов. Каждый шаблон является фрагментом кода (C++) и, возможно, содержит в себе специально отмеченные места (слоты) для вставки других шаблонов. Например, шаблон генетического алгоритма помимо кода самого алгоритма содержит слоты для вставки операторов отбора, скрещивания, мутации и т.д.

Набор шаблонов, которые могут быть вставлены в тот или иной слот (т.е. решающие одну и ту же задачу), образуют блок. Пользователь (на этапе диалога) должен выбрать один из имеющихся вариантов реализации тех блоков, которые входят в структуру рассматриваемого алгоритма. Блоки и входящие в них шаблоны зависят друг от друга (рис. 3). Имеется два вида такой зависимости. Во-первых, весь блок может напрямую зависеть от некоторого шаблона, если в этом шаблоне имеется слот для данного блока (например, блок оператора скрещивания зависит от шаблона генетического алгоритма и не зависит от шаблона метода роя частиц). Во-вторых, некоторые шаблоны

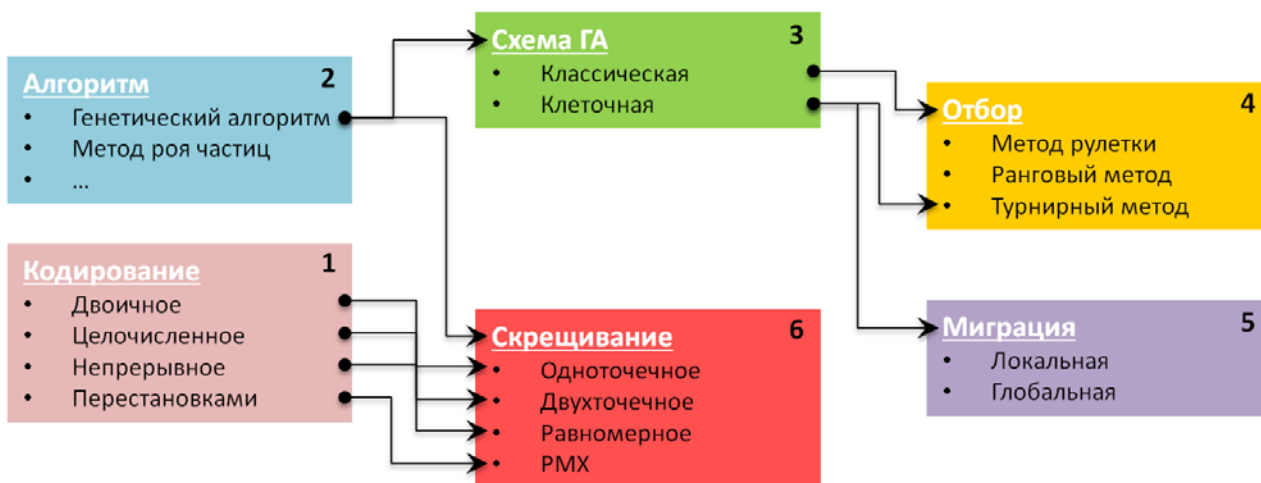


Рис. 3. Фрагмент конфигурационного файла системы

зависят от других шаблонов косвенным образом, если их работа должна быть согласована тем или иным способом. Например, если пользователем выбрана схема кодирования двоичными последовательностями, то для реализации оператора скрещивания будут доступны одноточечная, двухточечная и равномерная схемы. Если же было выбрано кодирование перестановками, то эти три схемы оказываются неприменимыми, а доступной для использования будет, например, схема PMX.

Вся информация о блоках и входящих в них шаблонах сохраняется в едином конфигурационном файле. Для каждого шаблона хранятся следующие данные:

- 1) название шаблона;
- 2) список шаблонов, от которых зависит данный шаблон;
- 3) список стандартных функций, вызовы которых имеются в данном шаблоне;
- 4) список числовых параметров, от которых зависит работа данного шаблона.

Важным является порядок хранения, потому что именно в этом порядке будет выполняться диалог с пользователем. Поэтому все блоки и шаблоны, от которых зависит данный шаблон, должны быть описаны раньше этого шаблона (порядок топологической сортировки).

Генерация кода сводится, таким образом, к постепенной детализации базового шаблона. В процессе этой детализации производится составление списка вспомогательных функций, на последнем этапе работы генератора код функций из этого списка вставляется в начало программы. Кроме того, составляется список всех числовых параметров, эти параметры организуются в специальную структуру, к которой имеется доступ из всех функций генерируемой программы.

В настоящее время разработан прототип системы, поддерживающий работу с генетическими и муравьиными алгоритмами, методом роя частиц,

алгоритмом бактериального поиска. Ведется работа по наполнению системы другими эволюционными алгоритмами. В будущем предполагается включить поддержку генерации параллельного кода (OpenMP, MPI, CUDA). Также планируется включить в систему базу тестовых задач для исследования эффективности разрабатываемых эволюционных алгоритмов оптимизации.

Работа выполнена при финансовой поддержке РФФИ (грант №14-07-00628 А).

## ЛИТЕРАТУРА

1. Ершов Н., Попова Н. Автоматическая генерация эволюционных алгоритмов под целевую архитектуру вычислительной системы // Тихоновские чтения. Научная конференция, Москва, факультет ВМК МГУ имени М.В. Ломоносова, 27-31 октября 2014 г. Тезисы докладов. — МАКС Пресс Москва, 2014. — С. 11.
2. D. Whitley, A Genetic Algorithm Tutorial, *Statistics and Computing* (4):65-85, 1994.
3. Ершов Н. Неоднородные клеточные генетические алгоритмы // Распределенные вычисления и ГРИД-технологии в науке и образовании: Тезисы докладов 6-й Международной конференции (Дубна, 30 июня - 5 июля 2014 г.). — ОИЯИ Дубна, 2014. — С. 68–69.
4. J. Kennedy, R. C. Eberhart, Particle swarm optimization, *Proceedings of IEEE International Conference on Neural Networks*, Piscataway, NJ, pp. 1942–1948, 1995.
5. Marco Dorigo and Thomas Stutzle // *Ant Colony Optimization*, The MIT Press, Cambridge, Massachusetts, 2004.

# БЕЗОПАСНОЕ SAAS ОБЛАКО ДЛЯ ПРОЕКТИРОВАНИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ ДОБЫЧИ НЕФТИ

А.А. Захаров<sup>1</sup> А.В. Бойко<sup>1</sup>, С.В. Степанов<sup>2</sup>

*1 Тюменский государственный университет*

*2 Тюменский нефтяной научный центр*

Задача увеличения извлечения нефти за счет новых моделей, методов и технологий была поставлена еще в "Основных направлениях экономического и социального развития СССР на 1981-1985 годы и на период до 1990 года". Решение задачи предусматривало в том числе создание банков гидрогеологической, геогидродинамической и гидрогеохимической информации, а также развитие технологий математического моделирования для оптимизации процессов разведки, добычи и транспортировки углеводородного сырья.

Отметим, что параллельно с созданием новых математических моделей развивается индустрия разработки программного обеспечения, позволяющая тиражировать отдельные прикладные программы на разные аппаратные платформы и создавать информационные и вычислительные системы "под ключ" путем интеграции различных сервисов в целевые проблемно-ориентированные комплексы. Практика решения задач, связанных с проектированием и сопровождением разработки месторождения нефти и газа показывает, что возможности существующего коммерческого программного обеспечения часто являются не достаточными. В этой связи создается оригинальное программное обеспечение, например, для оптимизации технологических режимов нефтяных горизонтальных скважин в подгазовых зонах [1], для анализа заводнения [2] и др.

Для ускорения внедрения новых программных разработок особую ценность представляет концепция облачных вычислений [3], в частности SaaS (Software as a service – прикладное ПО как услуга) [4] позволяющая создавать web- ориентированные лаборатории с интерактивным доступом и к инструментам моделирования и качественной поддержке пользователей: хранение исходных данных и результатов, визуализация процессов и т.п.

В рамках научного сотрудничества с Тюменским нефтяным научным центром авторами был выполнен пилотный проект по использованию безопасных облачных технологий для виртуальной лаборатории. В реализованном формате эта виртуальная лаборатория позволяет использовать комплекс оригинальных программ, в SaaS облаке. Эти программные сервисы имеют непосредственное отношение к анализу и проектированию разработки нефтяных и газонефтяных месторождений компаний «СургутНефтеГаз» и «Роснефть». Они позволяют исследовать причины сложной немонотонной динамики обводнения скважин, определять коэффициенты деления добычи и

закачки по пластам при их одновременной совместной разработке, а также прогнозировать технологический режим нефтяных горизонтальных скважин подгазовой зоны. Каждый из сервисов облачного программного комплекса также позволяет определять данные о системе «пласт-скважина», в том числе и в автоматическом режиме решения обратных задач [5-7]:

программа «*RecoveryDevisio*n» разделение добычи/закачки при одновременной совместной разработке пластов (значительное количество месторождений представлено именно многопластовыми залежами, разрабатываемых одновременно и совместно);

программа «*WellTuner*» исследование причин сложной немонотонной динамики обводнения вертикальной нефтяной скважины в условиях обводнения;

программа «*Sterkh*» имитация динамики газового фактора горизонтальной нефтяной скважины в условиях образования газового конуса.

Общим моментом в реализованных сервисах появляется возможность решать прямые и обратные задачи подземной гидродинамики применительно к конкретным производственным постановкам.

Пилотное внедрение комплекса показало, что одна из самых распространенных причин, по которым организации не спешат внедрять облачные ИТ сервисы, является безопасность исходных данных и результатов расчетов. Действительно, облачные вычисления не дают заказчику возможности контролировать не только технологические, но и собственные информационные ресурсы, а, следовательно, в такой области как подсчет запасов, где информация напрямую связана с налогообложением добывающих компаний, вопросы информационной безопасности выходят на первое место. Пользователи, посылая на обработку свои данные должны быть уверены в том, что их данные обрабатываются строго в соответствии с установленным технологическим процессом, учитывающим организационные и технические требования по обеспечению безопасности.

Применение ПО виртуализации как основы облачных технологий требует существенного изменения в подходах к обеспечению информационной безопасности систем [8]. Необходимо отметить появление нового, принципиально важного объекта виртуальной инфраструктуры – гипервизора, который на практике часто игнорируется и не защищается при помощи специализированных средств. Отметим, что только за счет компрометации гипервизора возможен вариант получения контроля над всеми подконтрольными ему виртуальными машинами и даже всей инфраструктурой виртуализации. В качестве защитных средств применялись: интеграция хост-серверов со службой каталога Active Directory, использование политик сложности и устаревания паролей, стандартизация процедур доступа к управляющим средствам хост-сервера, встроенный брандмауэр хоста виртуализации, применение единого профайла политик безопасности хоста, сторонние инструменты аудита и контроля доступа к подсистемам гипервизора.

Следует отметить, что при внедрении технологий виртуализации происходят серьезные изменения в физической инфраструктуре. С точки зрения организации сети возникает такое новое понятие как виртуальный коммутатор, который обеспечивает сетевое взаимодействие виртуальных машин в пределах одного хоста виртуализации. Проблема виртуальных коммутаторов заключается в не подконтрольности внутрисетевого трафика, а также в возможности прослушивания всего сетевого трафика между виртуальными машинами. Для решения проблемы прослушивания портов использован подход к организации сетей VLAN на базе виртуальных коммутаторов, где тегирование кадров происходит на уровне хоста виртуализации еще до попадания пакетов в физическую сеть.

Виртуальная машина является самым потенциально опасным объектом виртуальной инфраструктуры с точки зрения информационной защиты ввиду ее изначальной полной незащищенности и простоты модификации данных. Кроме того, такие технологии как «живая миграция» и «мгновенные снимки» способны послужить инструментом сокрытия следов присутствия в руках злоумышленника. В частности, проникнув в гостевую операционную систему виртуальной машины и имея достаточный контроль над системой управления хостом виртуализации, можно скрыть следы своего пребывания путем возврата к предыдущему снимку диска виртуальной машины (снапшоту).

Наконец, кража самих файлов мгновенных снимков виртуальной машины способна привести к серьезной утечке информации, поскольку они содержат в себе все последующие изменения данных на виртуальном диске и полный снимок оперативной памяти виртуальной машины с момента создания снимка. Исходя из выше сказанного, были выделены следующие основные типы угроз безопасности виртуальных сред:

- Атака на виртуальную машину;
  - из другой виртуальной машины;
  - на диск и файлы конфигурации виртуальной машины;
  - на сеть репликации виртуальных машин;
  - на сеть и систему хранения данных содержащей файлы виртуальной машины;
  - на средства резервного копирования виртуальной машины.
- Атака непосредственно на хост (гипервизор) виртуализации:
- из физической сети;
  - средствами скомпрометированного сервера управления виртуальной инфраструктурой;
  - через внутренние сервисы гипервизора SSH, WEB, TELNET и другие;
  - через агентов гипервизора от сторонних производителей;
- Атака на сервер управления виртуальной инфраструктурой:
- через гостевую ОС, обеспечивающую функционирование управляющих сервисов;
  - через СУБД сервера управления виртуальной инфраструктурой;

- через базу учетных записей;
- через сервис взаимодействия и мониторинга с хостами виртуализации;
- атака на ресурсы хоста виртуализации путем;
- неконтролируемого роста числа виртуальных машин;
- некорректного планирования разграничения пулов ресурсов;
- некорректного планирования растущих по мере заполнения виртуальных дисков VM;
- некорректного разграничения прав пользователей и групп виртуальной инфраструктуры.

Для решения наиболее острой проблемы – обеспечения защиты доступа к исходным данным и результатам расчетов, хранящихся на дисках виртуальных машин и подверженных описанным выше типам угроз, было успешно опробовано решение шифрования дисков в реальном времени и доверенной загрузки виртуальных машин на базе дополнительного программного комплекса SafeNet ProtectV. ProtectV представляет собой полное функциональное решение для шифрования данных в виртуализированных и облачных окружениях, которое позволяет осуществлять управление данными, администрирование данных и обеспечивает их прозрачность, а также помогает соблюдать требования информационной политики организации - клиента облачных ИТ услуг. С помощью системы шифрования ProtectV появилась возможность защитить критичные данные хранящиеся в виртуальных машинах конечных пользователей облака на протяжении всего их жизненного цикла: от момента подготовки и инициализации и до уничтожения.

Среди основных особенностей решения —гранулярное и полное шифрование всех виртуальных дисков виртуальной машины, поддерживает предстартовую аутентификацию и размещение отправной точки доверия на оборудовании на стороне заказчика, что обеспечивает комплексную защиту на всем протяжении жизненного цикла информации, позволяет запускать системы в многопользовательском окружении. Все виртуальные машины и соответствующие им разделы для хранения данных шифруются с применением надежного симметричного алгоритма блочного шифрования AES с длиной ключа 256 бит — сюда относятся копии дисков виртуальных машин (vmdk), их конфигурации (snapshot) и резервные копии на всех узлах и площадках аварийного восстановления.

Таким образом, привилегированные пользователи и администраторы облачной инфраструктуры, без прямой санкции владельца виртуальных машин, не могут получить доступ к зашифрованным виртуальным машинам. Для осуществления надлежащего контроля и обеспечения надежного управления аудитом вне зависимости от того, где размещаются или хранятся данные, соблюдая при этом требования законодательных нормативных актов (включая PCI DSS, HIPAA) используется механизм фиксируемого контроля с подтверждением операций по управлению данными через журналы аудита.

Для автоматизированного аудита виртуальной среды на предмет наличия ошибок в конфигурации безопасности виртуальной инфраструктуре VMware vSphere нами разработан программный продукт, который использует для взаимодействия с компонентами платформы виртуализации VMware vSphere стандартный VMware vSphere SDK интерфейс. На вход программе подается адрес конкретного хоста виртуализации VMware ESX либо сервера управления всей инфраструктурой VMware vCenter и учетные данные пользователя с правами на чтение. На выходе программа генерирует отчет по состоянию защиты исследуемого объекта и выставляет общий рейтинг защищенности на соответствие одному из 3-х уровней защищенности, предложенных компанией производителем VMware Inc:

1. Уровень предприятия (Enterprise) Этот уровень предназначен для защиты от большинства типичных атак на виртуальную инфраструктуру и обеспечения высокого уровня защищенности конфиденциальной информации.

2. Уровень демилитаризованной зоны (DMZ). Этот уровень позволяет обеспечить надежную защиту хостов и виртуальных машин, имеющих подключение к Интернет.

3. Уровень специализированной зоны с ограниченной функциональностью (SSLF). Этот уровень призван обеспечить максимально возможную степень защиты виртуальной инфраструктуры, в том числе за счет потери определенной функциональности виртуальной инфраструктуры в пользу защищенности от самых ухищренных атак.

Отчет представляет собой детализированную таблицу, разделенную по типам угроз, свойственных виртуальной инфраструктуре, которые были предложены выше. В качестве тестов на защищенность используется отслеживание параметров конфигурации хостов, виртуальных машин, сервера управления и 64 другие, основанные на рекомендуемых регламентах производителя платформы. В основе этих рекомендаций лежит технический документ VMware vSphere Hardening Guide, описывающий 3 уровня защищенности виртуальной инфраструктуры VMware vSphere, где каждому из этих уровней соответствует более 100 параметров объектов системы виртуализации. Все эти параметры аккумулируются и анализируются движком программы в автоматическом режиме и накладываются на заранее созданный шаблон угроз по уровню защищенности. В результате пользователь (администратор) может детально отследить, какому уровню защищенности соответствует данная виртуальная инфраструктура и на какие параметры системы следует обратить внимание для приведения ее в соответствие.

Предлагаемый программный комплекс в значительной мере повышает безопасность виртуальной инфраструктуры, однако, естественно, он не в состоянии обеспечить абсолютную защиту виртуальной среды. Следовательно, необходимо выработать и стандартизировать единый подход к обеспечению информационной безопасности в виде регламентов и стандартов, обязательно учитывая рекомендации производителя платформы виртуализации, поскольку



именно технологические особенности платформы определяют необходимые меры по обеспечению безопасности.

## ЛИТЕРАТУРА

1. Аре Мьеваттен, Роберт Осхайм, Стейнар Сэлид, Оддвар Груннинг. Модель образования газового конуса и зависимости газового фактора от темпа отбора в нефтеназоносном пласте с нефтяной оторочкой, SPE 102390, 2006.
2. M.Sayarpour, E.Zuluaga, C.S.Kabir, Larry W.Lake. The Use of Capacitance-Resistive Models for Rapid Estimation of Waterflood, SPE 110081, 2007.
3. Иванников В. П. Облачные вычисления в образовании, науке и госсекторе / В.П. Иванников // Параллельные вычисления и задачи управления : пленарные доклады V Междунар. конф. – М., 2010
4. Листопад Н. И., Олизарович Е. В. Модели функционирования "облачной" компьютерной системы. // Доклады БГУИР, №3(65) – 2012, с.23-29.
5. Степанов С.В. Численное исследование влияния капиллярного давления и сжимаемости на динамику обводненности скважины. // Нефтяное хозяйство, 08-2008, с. 72-74.
6. Степанов С.В., Степанов А.В., Елецкий С.В. Численно-аналитический подход к решению задачи оперативного прогнозирования работы нефтяной скважины в условиях образования газового конуса // Нефтепромысловое дело, 2/2013, с.53-58.
7. Степанов С.В., Гринченко В.А., Степанов А.В., Анурьев Д.А., Долгов И.А. Сопровождение разработки подгазовой зоны с использованием различных видов гидродинамического моделирования на примере Верхнечонского месторождения // Научно-технический вестник ОАО «НК «Роснефть», 4-2013, с.38-45.
8. А.А. Захаров, А.В. Бойко Безопасность НРС для интеллектуальных месторождений //Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (17-22 сентября 2012 г., г. Новороссийск). - М.: Изд-во МГУ, 2012. – с.63-65

# ИНДЕКСАЦИЯ ВИДЕОЗАПИСЕЙ НАУЧНЫХ МЕРОПРИЯТИЙ В ОБЩЕРОССИЙСКОМ МАТЕМАТИЧЕСКОМ ПОРТАЛЕ MATH-NET.RU

А.Д. Изаак, О.Г. Мисюрина, Ю.А. Пупырев, Д.Е. Чебуков

*Математический институт им. В.А. Стеклова РАН*

В базах данных Общероссийского математического портала Math-Net.Ru [1] наряду с научными журналами и статьями, опубликованными в них, индексируются научные мероприятия (конференции, семинары, летние школы, циклы лекций) и отдельные доклады, представленные на этих мероприятиях. К организации базы статей научных журналов и базы докладов научных конференций применяется одинаковый подход. Поддерживается интеграция между базами публикаций, докладов, персоналий и организаций.

Основные метаданные статьи в журнале и доклада на конференции приведены в таблице 1, в которой данные одного типа расположены на одной строке. Очевидно, что метаданные журнала в целом сходны с метаданными мероприятия: название журнала соответствует названию мероприятия, у журнала есть уникальный индекс ISSN, уникальность мероприятия обеспечивает время и место проведения, у журналов есть издатели, учредители, редколлегия, у конференций - организаторы и оргкомитет. Сравнивая статью и доклад, можно найти соответствующие аналогии между названиями доклада и статьи, авторами и их аффилиациями, аннотациями, ключевыми словами и списками литературы. Уникальность статьи обеспечивает год ее издания, том, выпуск, страницы и/или порядковый номер в выпуске, уникальность доклада - это дата и время и/или порядковый номер. Наконец, полному тексту статьи соответствует видеозапись выступления докладчика. Последние хранятся в виде файлов на локальном или удаленном сервере, и связаны с базой данных уникальным кодом.

<b>Статья в журнале</b>	<b>Доклад на конференции</b>
Название журнала	Название мероприятия
ISSN	Дата и место проведения
Издатель, учредитель	Организации
Редколлегия	Оргкомитет
Название статьи	Название доклада
Авторы (элементы БД Персоналии)	Авторы (элементы БД Персоналии)
Аффилиации (элементы БД Организации)	Аффилиации (элементы БД Организации)
Аннотация статьи	Аннотация (тезисы) доклада
Ключевые слова	Ключевые слова
Список литературы	Список литературы
Год, том, выпуск, страницы/номер	Дата, время/номер
Полный текст (PDF файл)	Видеозапись доклада (MP4 файл)

Табл. 1. Основные метаданные статьи в журнале и доклада на конференции.

Из статей автоматически формируются выпуски и тома журнала, из докладов аналогичным образом – программа конференции. На страницах статей и докладов присутствуют ссылки на персональные страницы авторов, так как авторы являются элементами базы данных Персоналии. На персональных страницах авторов автоматически собираются списки их публикаций в различных журналах, а также списки сделанных ими докладов на различных мероприятиях. На странице или сайте организации можно автоматически вывести как общий список публикаций сотрудников, так и список их докладов на различных научных мероприятиях. Помимо этого, можно организовать перекрестные ссылки между отдельными статьями (циклы публикаций), отдельными докладами (циклы докладов), а также между статьями и докладами сходной тематики.

Метаданные статьи хоть и могут дать читателям общее представление о теме работы, но являются лишь средством для поиска полного текста нужной статьи, информация о публикации без линка на полный ее текст не имеет смысла. Таким же образом метаданные доклада являются средством поиска его видеозаписи. Важнейшей частью нашей работы мы считаем не только индексацию докладов, но и организацию самого процесса видеосъемки научного мероприятия и монтажа снятого видео.

В Математическом институте им. В.А. Стеклова для видеосъемки научного мероприятия используются профессиональная видеокамера с возможностью записи на SD карты и входами для микрофонов, высококачественный штатив, петличный микрофон для докладчика (запись со встроенного микрофона камеры не допускается). Важнейшую роль играет работа оператора видеосъемки. Для организации прямой видео-трансляции камера должна быть также оборудована HDMI или SDI видео выходами для соединения с сервером трансляции. Монтаж производится в программе Adobe Premier (<http://www.adobe.com>), видео файлы сохраняются в формате MP4 в разрешениях FullHD 1920x1080, битрейт 5500 Mbps для просмотра на стационарных компьютерах с высокоскоростным подсоединением Internet и 720x480, битрейт 900 Mbps для просмотра на мобильных устройствах.

Видеофайлы хранятся на сервере, а доступ к ним осуществляется при помощи пакета LightHttpd с модулем H264 (в режиме *pseudo-streaming*) (<http://www.lighttpd.net>) или видео-сервера Wowza (режим *VOD: video on demand*) (<http://www.wowza.com>). В браузерах клиентов используется видео плеер JW Player (<http://www.jwplayer.com>). Данная конфигурация позволяет осуществлять быструю перемотку видео на любое время в обе стороны, что особенно важно при просмотре лекций и докладов длительностью более одного часа в формате FullHD, так как даже при высокоскоростном соединении с Internet видео файл такого размера (от 2Г до 4Г) не может быть загружен полностью достаточно быстро. Необходимо отметить, что видео файлы также можно загрузить на YouTube (<http://www.youtube.com>) и в метаданных доклада указать адрес соответствующей страницы, в этом случае вместо плеера JW

Player посетители увидят встроенный плеер YouTube с аналогичным функционалом.

Работа с метаданными осуществляется средствами системы Math-Net.Ru [2].

Просмотр видеофайлов возможен на устройствах, работающих на различных операционных системах и браузерах — в стационарных компьютерах под управлением Windows, Mac OS X или Linux просмотр осуществляется в формате Flash, на мобильных устройствах Apple и Android (в том числе мобильных телефонах) — в формате HTML 5. Переключение режима происходит автоматически при определении сервером названия браузера клиента.

Для каждого доклада производится регистрация статистики посещения страниц доклада и количества просмотров видеозаписей. Автоматически формируется список самых популярных мероприятий за неделю, за месяц и за все время. В таблице 2 приведены статистические данные о просмотрах видеозаписей некоторых научных мероприятий в 2014 г.: количество обработанных докладов, общее и среднее число просмотров этих материалов. Полный список видеозаписей расположен на странице "Видеотека Math-Net.Ru" <http://www.mathnet.ru/video>.

<b>Мероприятие</b>	<b>Видеозаписей</b>	<b>Просмотров</b>	<b>В среднем</b>
<i>Доклады на конференциях</i>	378	11561	30
<i>Лекции летней школы «Современная математика», г. Дубна</i>	152	8814	60
<i>Лекции и семинары Независимого московского университета</i>	123	10969	90
<i>Общеинститутские семинары Математического института им. В.А. Стеклова</i>	31	6420	210
<i>Популярные лекции по математике и смежным наукам</i>	6	2630	440
<i>Мемориальные конференции и воспоминания</i>	6	525	88

Таблица 2. Количество видеозаписей некоторых научных мероприятий и число их просмотров в 2014 г.

Наиболее популярными оказываются научно-популярные и учебные мероприятия, лекции учебных курсов для студентов и аспирантов, летние математические школы, а также общеинститутские семинары Математического института им. В.А.Стеклова. Количество просмотров значительно превышает число возможных участников мероприятия. При подсчете просмотров применяются рекомендации проекта COUNTER (<http://www.projectcounter.org>), что позволяет адекватно оценить число реальных просмотров посетителями, отфильтровать повторные просмотры и интернет-роботы.

Помимо видеосъемки мы организуем прямую трансляцию всех мероприятий, проходящих в актовом зале Математического института им. В.А.Стеклова - заседания Общеинститутских семинаров «Математика и ее приложения» и «Коллоквиум МИАН», пленарные доклады большинства конференций, отдельные лекции транслируются в прямом эфире на сайте МИАН. Трансляция прямого эфира ведется в форматах Flash, HTML5, SilverLight и RSTP, что позволяет просматривать видео на всех электронных устройствах, в том числе мобильных.

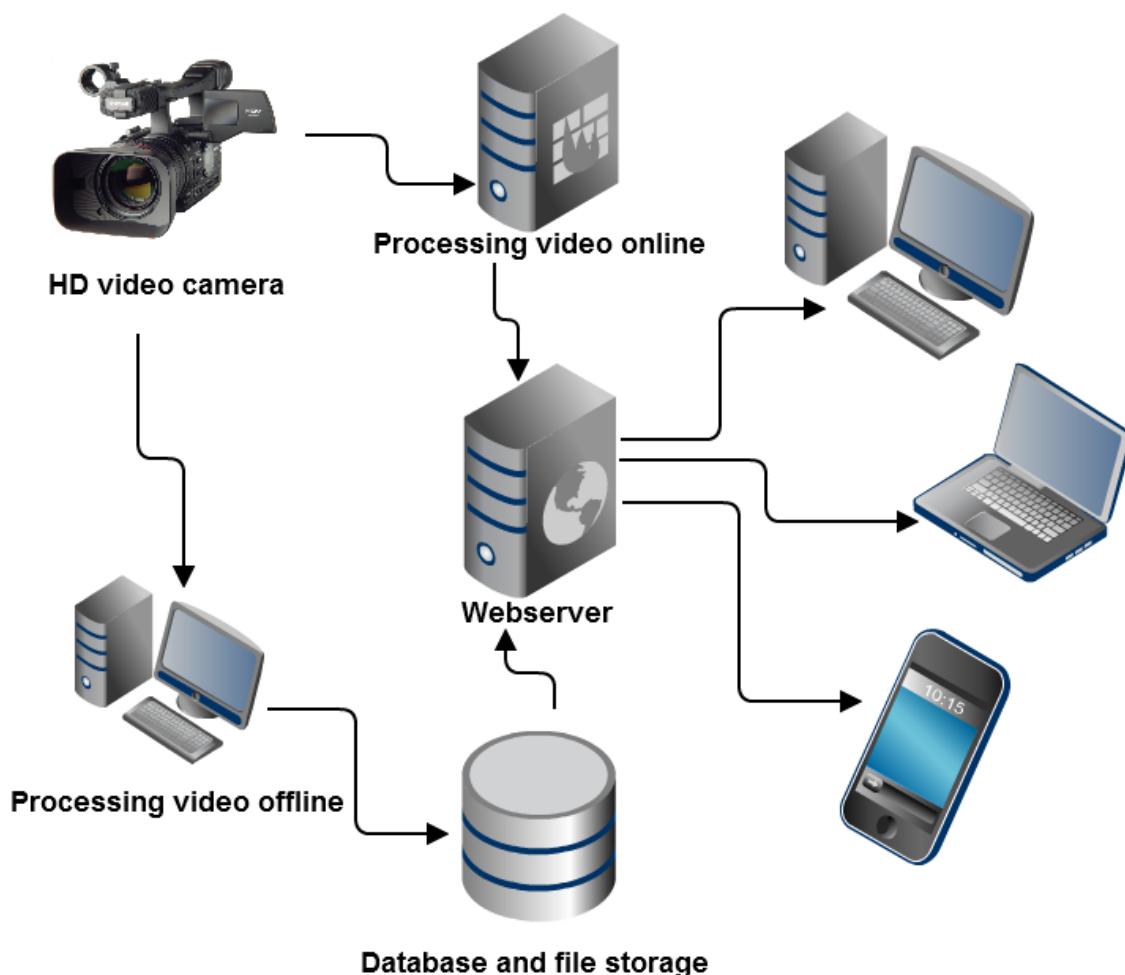


Рис. 1. Схема комплекса для организации видеосъемки, трансляции, монтажа и индексации видеозаписей научных мероприятий

Для организации прямой трансляции камера через HDMI или SDI видео выход подсоединяется к компьютеру, оборудованному картой видеозахвата Blackmagic (<https://www.blackmagicdesign.com>) и программным обеспечением, осуществляющим кодирование видео потока в формат H.264, и затем передающим сигнал на сервер Wowza. Сервер Wowza осуществляет трансляцию живого видео во всех необходимых форматах.

Таким образом, в Математическом институте им. В.А. Стеклова разработан и применяется полный комплекс для организации видеосъемки, трансляции, монтажа, индексации и представления в сети InterNet видеозаписей научных мероприятий. На рис. 1 приведена общая схема комплекса, включающая в себя описанные выше этапы видеосъемки, прямой трансляции, пост-обработки видео файлов, загрузки файлов на веб-сервер и просмотра видео на клиентских машинах.

Работа выполнена при частичной финансовой поддержке РФФИ (проект 13-07-00672 А).

## ЛИТЕРАТУРА

1. D. Chebukov, A. Izaak, O. Misyurina, Yu. Pupyrev, A. Zhizhchenko, “Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today”, *Intelligent Computer Mathematics*, Lecture Notes in Computer Science, **7961**, ed. J. Carette et al., Springer, 2013, 344–348.
2. Д. Е. Чебуков, А. Д. Изаак, О. Г. Мисюрина, Ю. А. Пупырев, “Электронный документооборот в информационной системе Math-Net.Ru как основа электронного издательства научного журнала”, *Научный сервис в сети Интернет: поиск новых решений*, Труды Международной суперкомпьютерной конференции (17–22 сентября 2012 г., г. Новороссийск), 2012.

# АНАЛИЗ ХАРАКТЕРИСТИК РАСПАДА АТОМНЫХ ЯДЕР В РАМКАХ СЕТЕВОЙ БАЗЫ ЗНАНИЙ NRV ПО НИЗКОЭНЕРГЕТИЧЕСКОЙ ЯДЕРНОЙ ФИЗИКЕ

А.В. Карпов<sup>1</sup>, А.С. Деникин<sup>1,2</sup>, А.П. Алексеев<sup>3</sup>, В.В. Самарин<sup>1</sup>,  
М.А. Науменко<sup>1</sup>, В.А. Рачков<sup>1</sup>

*1 Лаборатория ядерных реакций, ОИЯИ, Дубна, Россия*

*2 Международный университет “Дубна”, Дубна, Россия*

*3 Чувашский государственный университет, Чебоксары, Россия*

## ***1 Введение***

Сетевая база знаний NRV по низкоэнергетической ядерной физике [1-4] разработана в ЛЯР ОИЯИ, чтобы обеспечить быстрый доступ к актуальным экспериментальным данным по ядерной структуре и сечениям ядерных реакций, а также для анализа данных и моделирования процессов ядерной динамики в рамках устоявшихся физических подходов.

Существуют три принципиальных отличия создаваемой нами ядерной базы знаний от уже имеющихся ядерных баз данных:

1) Известные нам базы ядерных данных, как правило, предоставляют пользователю обычные текстовые файлы, позволяющие получать информацию лишь в табличной форме. Таким образом, даже для получения каких-либо простейших систематик пользователь вынужден самостоятельно выбирать и переписывать в отдельный файл нужные ему данные, а затем использовать какой-либо графический пакет для их представления. Наша база данных содержит специальные программы их графического представления, проведения их сравнительного анализа и получения всевозможных систематик по группе или по всей карте ядер.

2) Организованные нами базы данных по сечениям ядерных реакций содержат оцифрованные данные. Это не только облегчает их использование, но, прежде всего, позволяет с помощью разработанных нами программ выполнять всевозможную обработку этих данных, делать графическое сравнение различных сечений друг с другом и, в конечном счете, проводить полный теоретический анализ этих данных. Все это выполняется непосредственно в окне веб-браузера, без необходимости поиска и загрузки каких-либо дополнительных вычислительных программ и графических пакетов.

3) Главной частью разрабатываемого нами комплекса являются сложные вычислительные программы моделирования низкоэнергетической ядерной динамики, которые в совокупности с экспериментальными ядерными данными как раз и составляют “ядерную базу знаний”. Организация сетевой работы этих программ (графический интерфейс, простота в использовании, разветвленное меню, гипертекстовая система подсказок и описания моделей, графическое

представление результатов, взаимодействие с удаленным пользователем при длительных вычислениях и, наконец, многопользовательский сетевой доступ) представляет основную проблему создания базы знаний данного типа.

Сетевая база знаний NRV содержит большинство имеющихся экспериментальных данных о свойствах ядер, а также данные о сечениях различных ядерных реакций, включая слияние, образование остатков испарения и упругое рассеяние.

Доступные коды включают интерактивную карту ядер, оболочечную модель ядра, оптическую модель упругого рассеяния, метод связанных каналов для описания слияния ядер, статистическую модель распада ядра, коды для анализа неупругого рассеяния ядер и нуклонных передач (DWBA, FRESCO), код для кинематического анализа ядерных реакций с 2-х и 3-х частичным выходным каналом и многие другие.

В данной работе приводится краткое описание моделей распада атомных ядер, реализованных в базе знаний NRV, а также особенностей их реализации. На данный момент доступны модели для оценки свойств радиоактивного распада (альфа-распад, бета-распад, спонтанное деление), а также модели для оценки свойств вынужденного деления и распада возбужденных ядер.

## 2 Альфа-распад

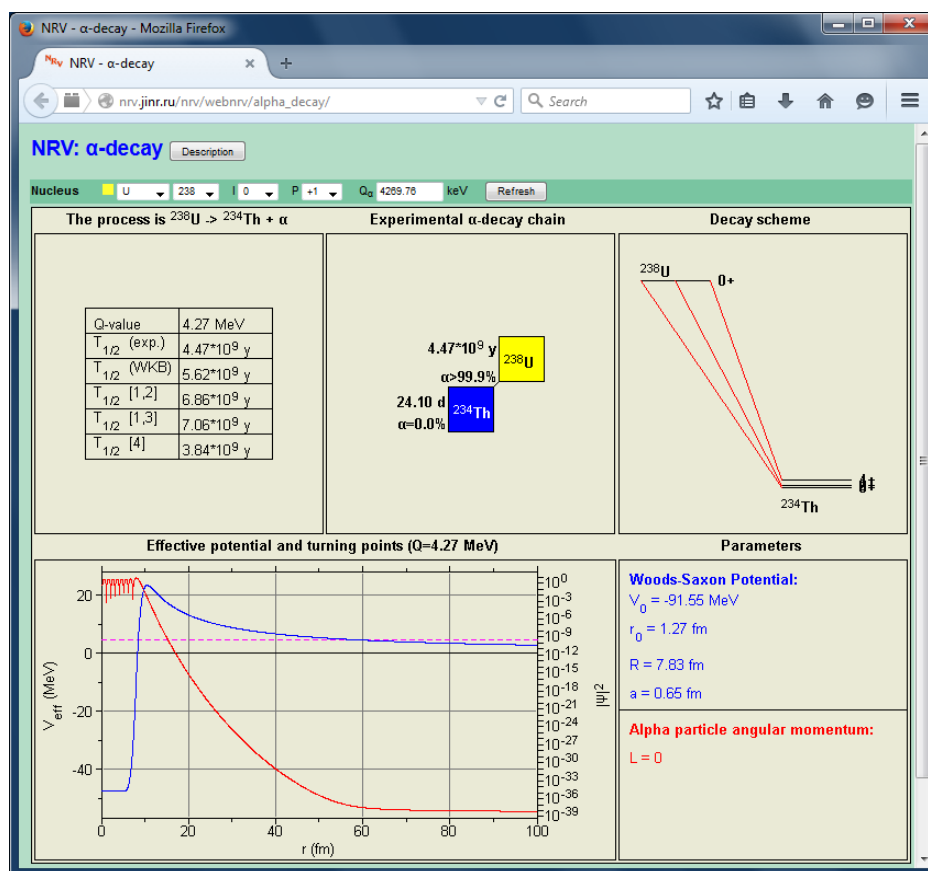


Рис. 1. Альфа-распад (веб-страница)



Данный раздел проекта NRV позволяет изучать свойства процесса альфа-распада: период полураспада, экспериментальную цепочку распада, схему распада, а также вид волновой функции, рассчитанной для определенного потенциала, параметры которого можно варьировать, используя соответствующий интерфейс (рис. 1). Код основан на работах [5-11]. Для вычисления энергии распада используются экспериментальные массы основных состояний, если они доступны в базе данных NRV, в противном случае используется теоретическая оценка [5]. Спины и четности основного состояния также берутся из базы данных, если они доступны, в противном случае предполагаются минимально возможные спины и положительная четность. Для оценки периода полураспада используются ВКБ приближение, а также эмпирические соотношения Виолы-Сиборга [6] с параметрами [7], соотношение Виолы-Сиборга с параметрами [8], а также соотношение Собичевского и др. [9]. Имеется подробное описание модели, которое можно увидеть при нажатии кнопки “Description”.

### 3 Бета-распад

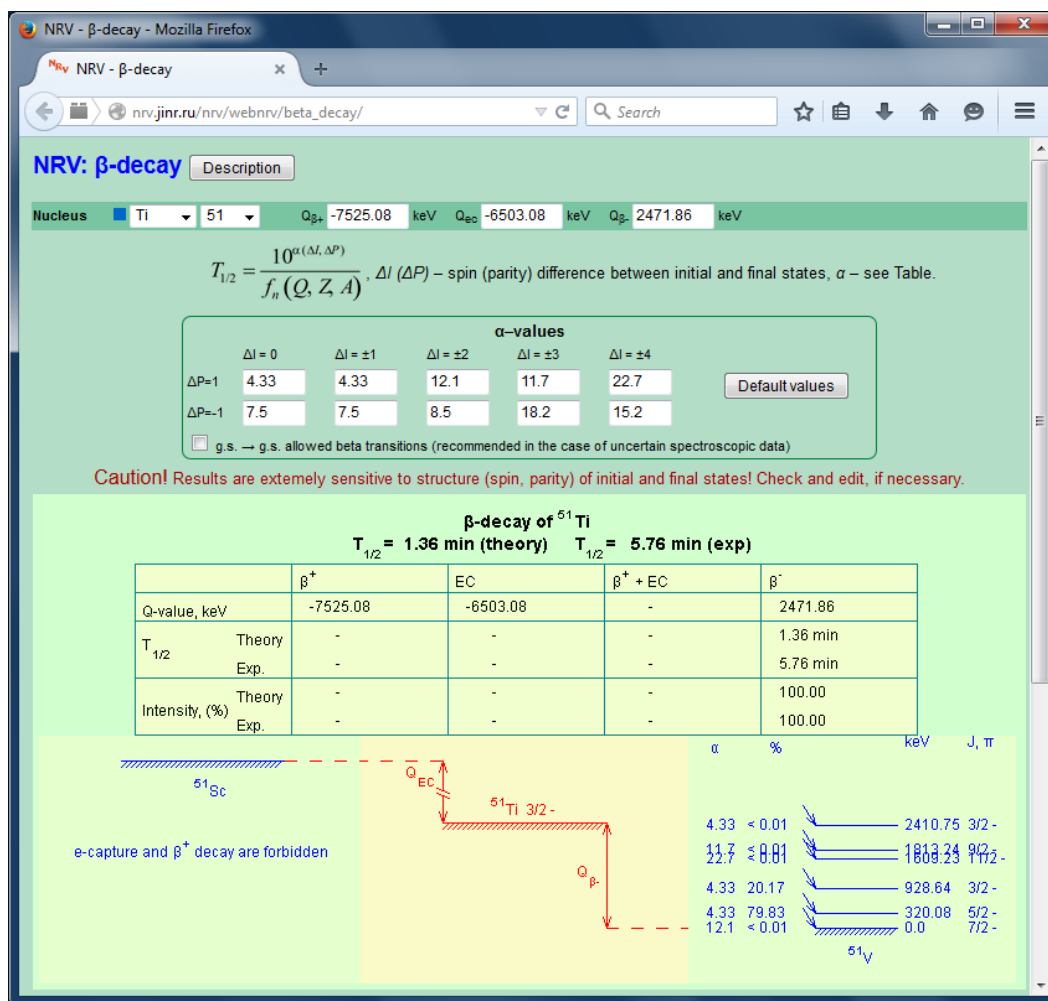


Рис. 2. Бета-распад (веб-страница)

Данный раздел проекта NRV позволяет изучать свойства процесса бета-распада: разрешенные виды распадов, период полураспада и схему распада (рис. 2). Код основан на работах [12-14]. Полный период полураспада рассчитывается как  $1/T_{1/2}^b = 1/T_{1/2}^{\beta^-} + 1/T_{1/2}^{\beta^+} + 1/T_{1/2}^{EC}$ , где каждое слагаемое это сумма по всем энергетически разрешенным состояниям  $E_f$  дочернего ядра

$$T_{1/2}^b = \left( \sum_{E_f < Q_b} \frac{10^{\alpha(\Delta I, \Delta P)}}{f_n(Q, Z, A)} \right)^{-1}, \quad b = \beta^+, \beta^-, EC.$$

Значения  $\alpha$  можно варьировать (по умолчанию они берутся из [13]), вид функции  $f_n(Q, Z, A)$  зависит от типа распада и степени разрешенности соответствующего перехода, которая в свою очередь определяется разностью спинов  $\Delta I$  и разностью четностей  $\Delta P$  между начальным и конечным состояниями. В случае отсутствия надёжных спектроскопических данных можно ограничиться выбором переходов только в основное состояние дочерних ядер (это осуществляется установкой соответствующей галочки). Энергии, спины и четности состояний берутся из нашей базы данных, но при этом также предусмотрена возможность варьировать их вручную. Имеется подробное описание модели, которое можно увидеть при нажатии кнопки “Description”.

#### 4 Деление

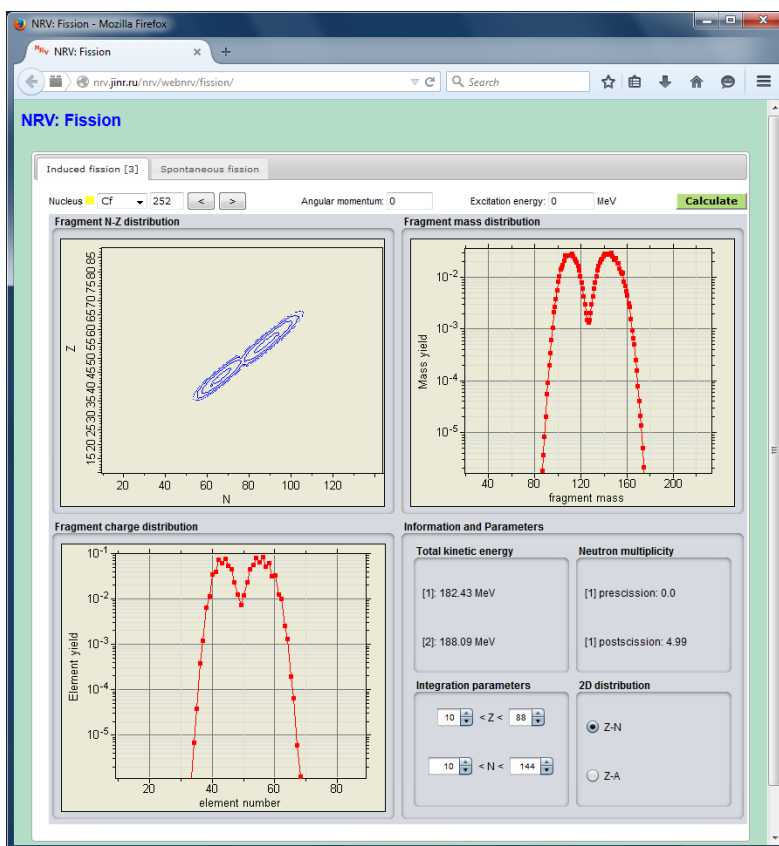


Рис. 3. Вынужденное деление (веб-страница)

Данный раздел проекта NRV позволяет изучать свойства процесса вынужденного деления и спонтанного деления.

Для вынужденного деления (рис. 3) выводится  $Z-N$  (либо  $Z-A$ ) распределение фрагментов, массовое распределение и зарядовое распределение. В расчётах указанных распределений используется код GEF [15], специально адаптированный для работы в составе базы знаний NRV. Множественности пред-делительных и пост-делительных нейтронов рассчитываются согласно [16]. Полная кинетическая энергия осколков рассчитывается согласно [16,17]. Веб-интерфейс позволяет варьировать энергию возбуждения и угловой момент ядра, а также диапазон учитываемых  $Z, N, A$ .

Для спонтанного деления (рис. 4) выводятся экспериментальный период полураспада, взятый из нашей базы данных (при наличии), а также теоретические оценки, рассчитанные согласно [14,18,19].

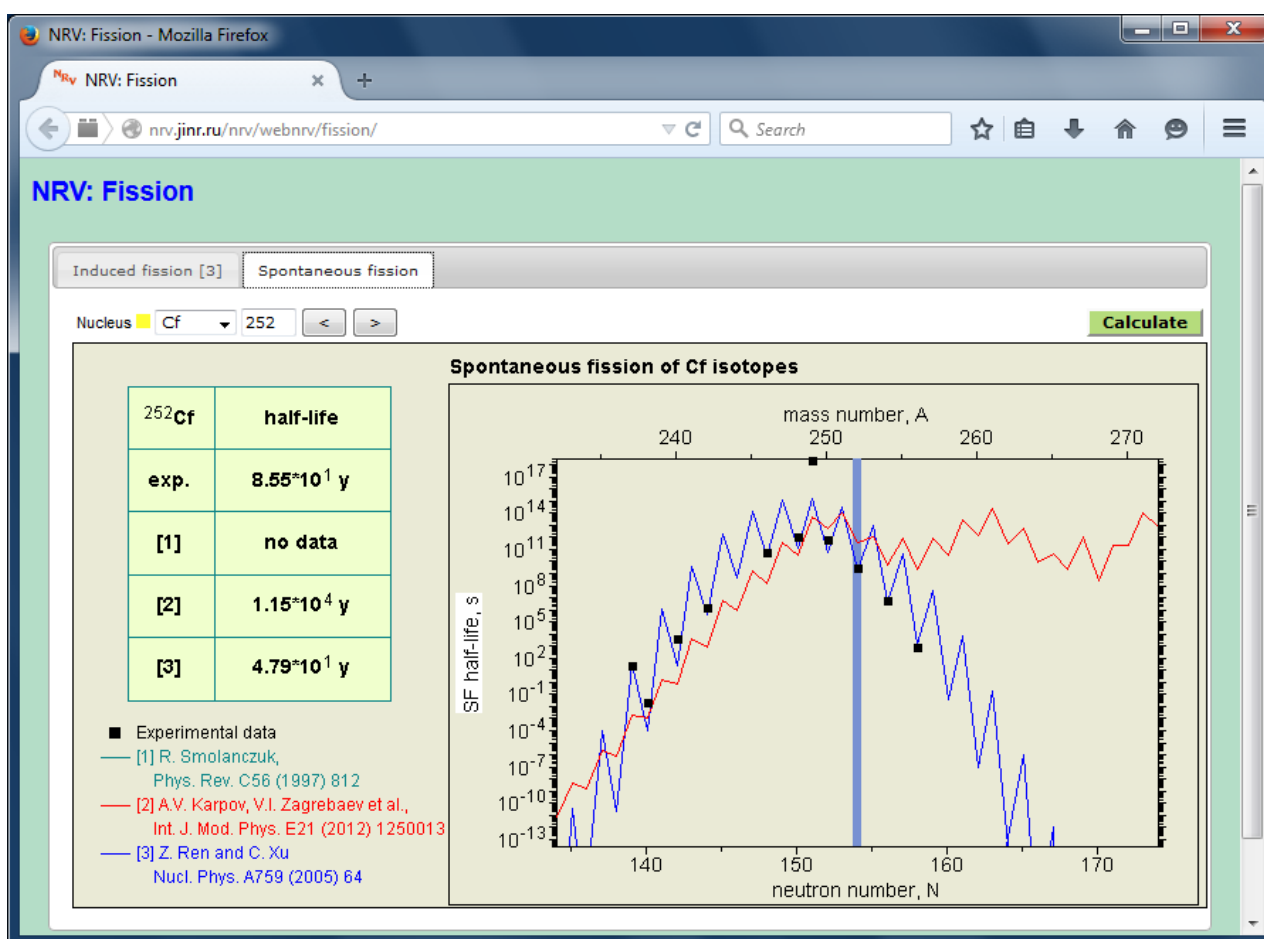


Рис. 4. Спонтанное деление (веб-страница)

## 5 Распад возбужденных ядер

Данный раздел проекта NRV позволяет рассчитывать ширины распада, а также вероятности выживания возбужденных ядер (рис. 5). В основе кода лежит статистическая модель распада возбужденного вращающегося ядра [5,20-21]. Свойства ядер берутся из нашей базы данных (при наличии), но при этом также предусмотрена возможность варьировать их вручную. Кроме того, интерфейс предоставляет возможность варьировать диапазон энергий возбуждения, параметры плотности уровней, момент инерции, вид и параметры коллективного усиления плотности уровней. Имеется подробное описание модели, которое можно увидеть при нажатии кнопки “Description”.

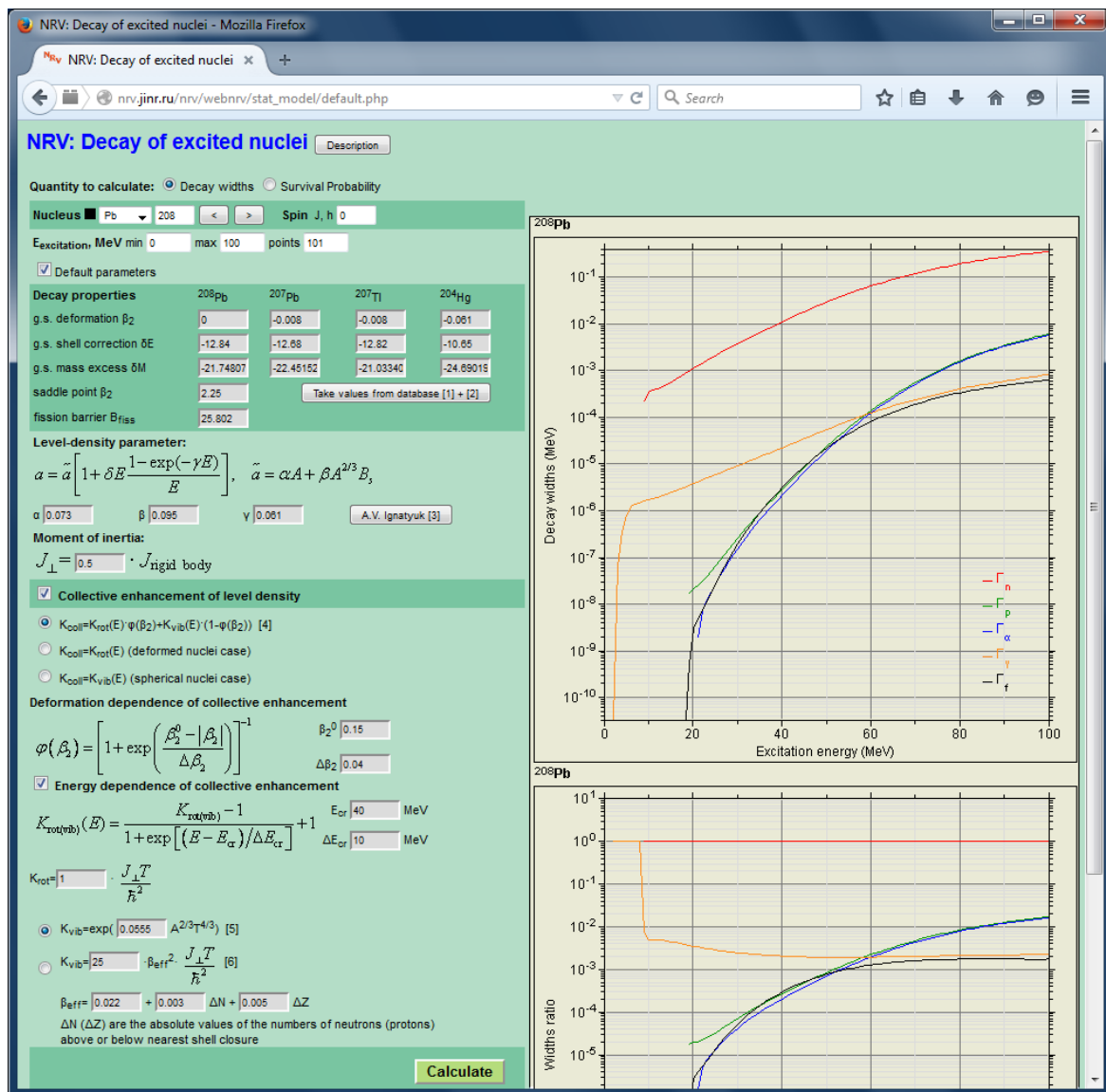


Рис. 5. Распад возбужденных ядер (веб-страница)

## ***6 Инструменты разработки***

При разработке базы знаний использованы MySQL, PHP, Java, JavaScript. Были созданы различные Java-апплеты, предназначенные для построения графиков и использующиеся практически во всех разделах базы знаний. Они позволяют не только гибко настраивать параметры отображения (масштаб, диапазон значений, отображение грид-линий и т.п.), но и сохранять данные в форматах ASCII, EPS, GIF, JPG и PNG. К сожалению, в последние несколько лет браузеры блокируют Java-апплеты по соображениям безопасности, и настройка браузера для корректной работы с Java становится нетривиальной задачей. В связи с этим было принято решение о переходе к использованию JavaScript для построения графиков на элементе Canvas, определяемом в стандарте HTML5. На данный момент разработан JavaScript-код для отображения статических (не интерактивных) графиков, который уже используется в некоторых разделах базы знаний. В будущем планируется полностью отказаться от использования Java, что потребует довольно много усилий и является частью наших ближайших планов.

## ***7 Заключение***

Разработанная сетевая база знаний NRV по низкоэнергетической ядерной физике обеспечивает быстрый доступ к современным экспериментальным данным по ядерной структуре и сечениям ядерных реакций, а также предоставляет возможность анализа данных и моделирования процессов ядерной динамики в рамках устоявшихся физических подходов.

В данной работе приведено краткое описание моделей распада атомных ядер, реализованных в базе знаний NRV, а также особенностей их реализации.

Сетевая база знаний NRV имеет несколько уникальных преимуществ по сравнению с другими ядерными базами данных и в настоящее время широко используется не только для научных целей, но также в качестве ценного инструмента в процессе образования в области ядерной физики.

Работа поддержана грантом 15-07-07673-а Российского фонда фундаментальных исследований (РФФИ).

## **ЛИТЕРАТУРА**

1. Сетевая база знаний NRV по низкоэнергетической ядерной физике, <http://nrv.jinr.ru/>
2. А.С. Деникин, А.П. Алексеев, В.И. Загребаев, М.А. Науменко, В.В. Самарин. “Сетевая база знаний по ядерной физике низких и промежуточных энергий”. Научный сервис в сети Интернет: технологии параллельного программирования. Труды Всероссийской научной конференции, Новороссийск, 2006, Изд-во МГУ, 2006, 211.
3. А.С. Деникин, А.В. Карпов, А.П. Алексеев, В.И. Загребаев, М.А. Науменко, В.В. Самарин. “Сетевая ядерно-физическая база знаний”. Научный сервис в сети Интернет: решение больших задач. Труды

- Всероссийской научной конференции, Новороссийск, 2008, Изд-во МГУ, 2008, 393.
4. A.S. Denikin, V.I. Zagrebaev, A.V. Karpov, A.P. Alekseev, N.M. Jacobs, T.S. Maluleke, Web knowledge base on low-energy nuclear physics, Proc. of 2nd South Africa - JINR Symposium: Models and Methods in Few- and Many-Body Systems, Dubna, Russia (2010).
  5. P. Möller, J.R. Nix, W.D. Myers et al., Atomic Data and Nuclear Data Tables 59, 185 (1995).
  6. V.E. Viola and G.T. Seaborg, J. Inorg. Nucl. Chem. 28, 741 (1966).
  7. Yu.Ts. Oganessian, V.K. Utyonkov, Yu.V. Lobanov et al., Phys. Rev. C 70, 064609 (2004).
  8. A.Sobiczewski, Z. Patyk, and S. Cwoik, Phys. Lett. B 224, 1 (1989).
  9. A.Parkhomenko, and A. Sobiczewski, Acta Phys. Pol. B 36, 3095 (2005).
  10. M.Wang, G.Audi, A.H.Wapstra et al., Chinese Physics C 36, 1603 (2012).
  11. В.В.Самарин. Изв. РАН Сер. физ. 78, 1388 (2014).
  12. Б.С. Джелепов, Л.Н. Зырянова, Ю.П. Суслов, “Бета-процессы. Функции для анализа бета-спектров и электронного захвата”. (Л.: “Наука”, 1972).
  13. C.S. Wu and S.A. Moszkowski, “Beta decay” (John Wiley & Sons, New York, 1966).
  14. A.V. Karpov, V.I. Zagrebaev et al. Int. J. Mod. Phys. E 21, 1250013 (2012).
  15. B. Jurado, K.-H. Schmidt, <http://www.khs-erzhausen.de/GEF.html>
  16. M.G. Itkis and A.Ya. Rusanov, Physics of Particle and Nuclei 29, 160 (1998).
  17. V.E. Viola, K. Kwiatkowski, M. Walker, Physical Review C 31, 1550 (1985).
  18. R. Smolanczuk. Phys. Rev. C 56, 812 (1997).
  19. Z. Ren and C. Xu, Nucl. Phys. A 759, 64 (2005).
  20. А.В. Игнатьюк, “Статистические свойства возбужденных атомных ядер”. (М.: Энергоатомиздат, 1983).
  21. V.I. Zagrebaev, Y. Aritomo, M.G. Itkis, Yu.Ts. Oganessian, M. Ohta, Phys. Rev. C 65, 014607 (2001).

# ИСПОЛЬЗОВАНИЕ ПРОГРАММ АССОЦИАТИВНОГО КАРТИРОВАНИЯ ДЛЯ УПРАВЛЕНИЯ РАСПРЕДЕЛЁННЫМИ ИНФОРМАЦИОННЫМИ РЕСУРСАМИ

Ф.О. Каспаринский<sup>1,2</sup>

*1 Биологический факультет Московского государственного университета*

*имени М.В.Ломоносова*

*2 ООО «МАСТЕР-МУЛЬТИМЕДИА»*

Совершенствование сетевого инструментария информационных технологий в период с 2010 по 2015 год привело к реформированию способов получения, хранения, обработки и использования данных, информации и медиаресурсов [1]. Утилитарные облачные сервисы [2] предоставили возможность перераспределять и синхронизовать информационные ресурсы между локальными и облачными (удалёнными сетевыми) хранилищами. Полиморфизм десктопных и мобильных программно-аппаратных платформ с сопутствующими облачными сервисами (Windows и OneDrive, Apple и iCloud, Android и Google Drive), независимых неспециализированных файловых хостингов (Dropbox, Vitcasa, Яндекс.Диск и др.) и медиахостингов (Vimeo, Flickr, MediaFire и др.), а также социальных сетей (Facebook, ВКонтакте, МойМир и др.) обострил проблему управления разобщенными информационными ресурсами.

Эволюция средств визуализации управления сложными наборами данных привела к возникновению применяемых в проектном менеджменте диаграмм Гантта [3] и систематизирующих знания ассоциативных карт Бьюзена [4]. В диаграммах Гантта входящие в проект фазы, стадии и этапы с их элементами иерархически размещаются по вертикали, формируя по горизонтальной временной шкале нисходящую лестницу графических плашек (отрезков) с обозначенной степенью завершения работы. Современный инструментарий, применяющий диаграммы Гантта (Microsoft Project [5] и др.), специализирован для использования в бизнесе: он позволяет определить критические стадии процесса, рационально перераспределить ресурсы и аргументировать материальное вознаграждение работников. Однако диаграммы Гантта не востребованы для менеджмента научной работы, поскольку не обладают средствами визуализации творческого процесса работы со знаниями.

Ассоциативные карты (mind maps, интеллект-карты) наглядно представляют общее системное мышление посредством построения диаграмм связей центрального элемента (идеи, темы, задачи, процесса) с древовидными группами дочерних элементов, градиент иерархического положения которых символизирует удаленность от центра и радиальное расположение по часовой стрелке. Иерархическая нумерация динамически меняется в соответствии с

положением элемента в системе. Элементы индивидуализируются шрифтовыми и цветовыми выделениями, маркировкой ярлычками-пиктограммами с общепринятыми одиночными (информационные, финансовые, офисные знаки) и серийными символами (номер, приоритет, степень завершения, смайлики, флаги, стрелки, булавки, звезды) или специальными анонсовыми образами. Односторонние и двухсторонние связи между любыми удаленными элементами позволяют поддерживать между ними ассоциативные связи. Возможно образование многоядерных структур и последующее установление связей между ними [6]. Элементы ассоциативных карт можно объединять общими границами с различным дизайном и впоследствии использовать особенности оформления для быстрого переключения внимания.

Утилитарность механизмов ассоциации элементов с текстовыми, табличными, графическими и аудиовизуальными данными, гиперссылками и файлами предопределяет возможность использования ассоциативных карт для создания наглядной системы знаний, протоколирования экспериментов, подготовки публикаций и выступлений. За 40 лет практического использования сформировался основной набор способов применения ассоциативных карт: восприятие информации (обучение), преподнесение информации (преподавание), планирование способов решения задач, мозговой штурм (генерация новых знаний) и аргументация принятия решений. В менеджменте научной работы ассоциативные карты в своем исходном виде сложно использовать, поскольку они не обеспечивают временной контроль хода выполнения задач.

Симбиотическое слияние диаграмм Гантта и ассоциативных карт в едином программном инструменте адаптировало их для применения в повседневной практике научной работы [6]. Среди лучших примеров – разработанная компанией Тони Бьюзена программа iMindMap [7], графический органайзер ConceptDraw MINDMAP [8] от CSO Corp., менеджер малых бизнес-проектов и учебной деятельности MindGenius [9] компании MindGenius Ltd., полиморфная интеллект-карта для научно-образовательной сферы и бизнеса MindView [10] компании MatchWare A/S, интегрированный с Microsoft SharePoint [11] администратор крупных проектов MindManager [12] компании Mindjet, оптимизированный для аналитиков MindMapper [13] компании SimTech Systems, а также PersonalBrain и BrainEKP [14] компании TheBrain Technologies LP, способные визуализировать данные в виде трехмерного облака с автоматическим сворачиванием и разворачиванием ветвей в каждом из узлов в зависимости от положения курсора. Каждая программа традиционно представлена несколькими вариантами, отличающимися набором функциональных возможностей и ценой владения.

В результате тщательного сопоставления возможностей вышеперечисленных программ в качестве инструмента научной работы в 2009 году мы выбрали наиболее продвинутый вариант MindView 3 BE с



возможностью мультиформатного импорта и экспорта информации (Word, PowerPoint, Excel, Outlook, Project, HTML, RTF, PDF, XML, Picture) и шестью альтернативными способами представления сведений (радиантная ассоциативная карта, иерархические древовидные диаграммы по вертикали и горизонтали, временная шкала, таблица и диаграмма Ганта) [6]. В ходе пятилетней работы было создано 75 ассоциативных карт, которые использовались для организации экспериментальной научной работы и бизнеса; управления односайтовыми и многосайтовыми интернет-проектами, подготовки статей и книг, иллюстрирования объёмных изданий, подготовки публичных выступлений и преподавания.

Эмпирическим путем было обнаружено, что используемая в программе MindView система лицензирования с онлайн-активацией и привязкой к аппаратной части компьютера отключает редакторский функционал программы при обновлении и расширении компьютерного парка. Просмотр ассоциативных карт MindView без покупки дополнительных лицензий возможен только после установки специальной программы-вьюера без редакторских опций. Вся информация о структуре ассоциативной карты MindView хранится в одном управляющем файле, доступ к которому существенно замедляется по мере увеличения количества элементов и ассоциированных с ними внутренних информационных ресурсов (штатные пиктограммы, тексты с простым форматированием, таблицы и гиперссылки). Непосредственная вставка копий интернет-страниц в карту MindView невозможна. В отличие от таблиц и текста, графическая и аудиовизуальная информация пользователя не включается в управляющий файл, а связывается с элементами карты посредством прямых ссылок на соответствующие независимые файлы. Просмотр и вызов связанных с элементом внешних источников информации возможен только после открытия выпадающего списка в результате нажатия на значок прикрепленной к элементу «скрепки», что снижает оперативность доступа к документам. Было установлено, что существующий функционал программы MindView позволяет эффективно управлять информационными массивами, содержащими до 1000 элементов, связанных иерархическими связями с одним центром. Автоматическая оптимизация позиционирования элементов не рассчитана на установление соседских связей между элементами или центральными ядрами и создает контрпродуктивное визуальное нагромождение. Результаты поисковых запросов в ассоциативной карте MindView предъявляются последовательно посредством графического выделения соответствующего элемента, что замедляет работу при большой поисковой выдаче. Синхронизация содержимого управляющих файлов MindView 2009 г. выпуска не была предусмотрена. Несмотря на прогресс облачных технологий [2] за последние 5 лет большинство вышеперечисленных программ ассоциативного картирования не обогатилось утилитарными инструментами для коллективной работы с разнородными распределенными информационными ресурсами: для обмена информацией ассоциативных карт между удаленными пользователями теперь

предлагается упаковка связанных ресурсов в пересылаемый архив или приобретение программных надстроек для синхронизации через сеть в реальном времени.

Программа PersonalBrain трансформировалась в утилитарный TheBrain [14] после избавления от когнитивно бесполезного трехмерного представления элементов [15]. TheBrain отличается от прочих программ ассоциативного картирования автоматическим центрированием в рабочем пространстве любого выделенного элемента с оптимальным перераспределением позиций множественных связанных родительских, соседских и дочерних элементов. Связи между элементами современной карты TheBrain 8 визуализуются экспоненциальными кривыми, которые могут выделяться цветом, толщиной, фактурой и текстовой меткой, а также иметь одностороннюю направленность. Если количество связанных с узлом элементов одного уровня иерархии превосходит когнитивно комфортный предел, они самоорганизуются в списки с автоматическим скроллингом. Пользователь может варьировать отображаемое количество уровней ассоциативных узлов, связанных с центральным элементом или морфировать карту в иерархический список с произвольным началом и расположением элементов, после чего сохранить созданное визуальное представление для последующего использования (презентация, продолжение работы и т.п.). В специальном демонстрационном режиме карта автоматически переключает центр внимания между всеми разрешенными к показу элементами. Профессиональная версия программы TheBrain 8 Pro обеспечивает возможность группировки произвольных элементов посредством любого количества пользовательских тегов или одного из типов, который может входить как элемент в отдельную иерархическую структуру. Принадлежность к определенному типу автоматически предопределяет стиль названия элемента (шрифт, кегль, цвет и аносовое изображение).

Организация связей между элементами, в том числе скрытыми из поля зрения, облегчается интеллектуальной системой поиска, автоматически предлагающей возможные варианты по нескольким первым буквам любых слов из названия искомого элемента, отсортированные с учетом последней информации, введенной пользователем в любое компьютерное приложение посредством клавиатуры или буфера обмена. Исходя из этой особенности, названия элементов с первичными данными [1], которые впоследствии будут консолидироваться в информационные ресурсы [1], целесообразно строить по формуле «Когда Где Кто/Что Чем Кем», где «Когда» - обратная дата получения данных в формате ГГГГММДД; «Где» - место получения данных; «Кто/Что» - объект или тематика наблюдений; «Чем» - регистрирующая аппаратура; «Кем» - автор данных (оператор). При добавлении элементов и ассоциированных с ними внутренних и внешних информационных ресурсов, включая веб-страницы, система автоматически индексирует их содержимое для расширенного поиска, результаты которого выдаются списком и сохраняются для будущего использования.

Эффективность поисковой подсистемы TheBrain в совокупности с потенциальной возможностью множественного связывания одного информационного ресурса с несколькими категориями (директориями) позволяет использовать ассоциативную карту TheBrain в качестве менеджера локальных файлов. Для этого в программе предусмотрен автоматический импорт директорий (папок) и их файлов с установлением логических связей без ограничения уровней вложенности. Локальные или сетевые файлы любых типов могут быть связаны с элементами TheBrain посредством гиперссылок, создающихся методом «drag-and-drop» с автоматическим импортом анонсовых изображений, пиктограмм файловых типов или фавиконок веб-страниц, при активации которых происходит вызов приложения, соответствующего типу файла.

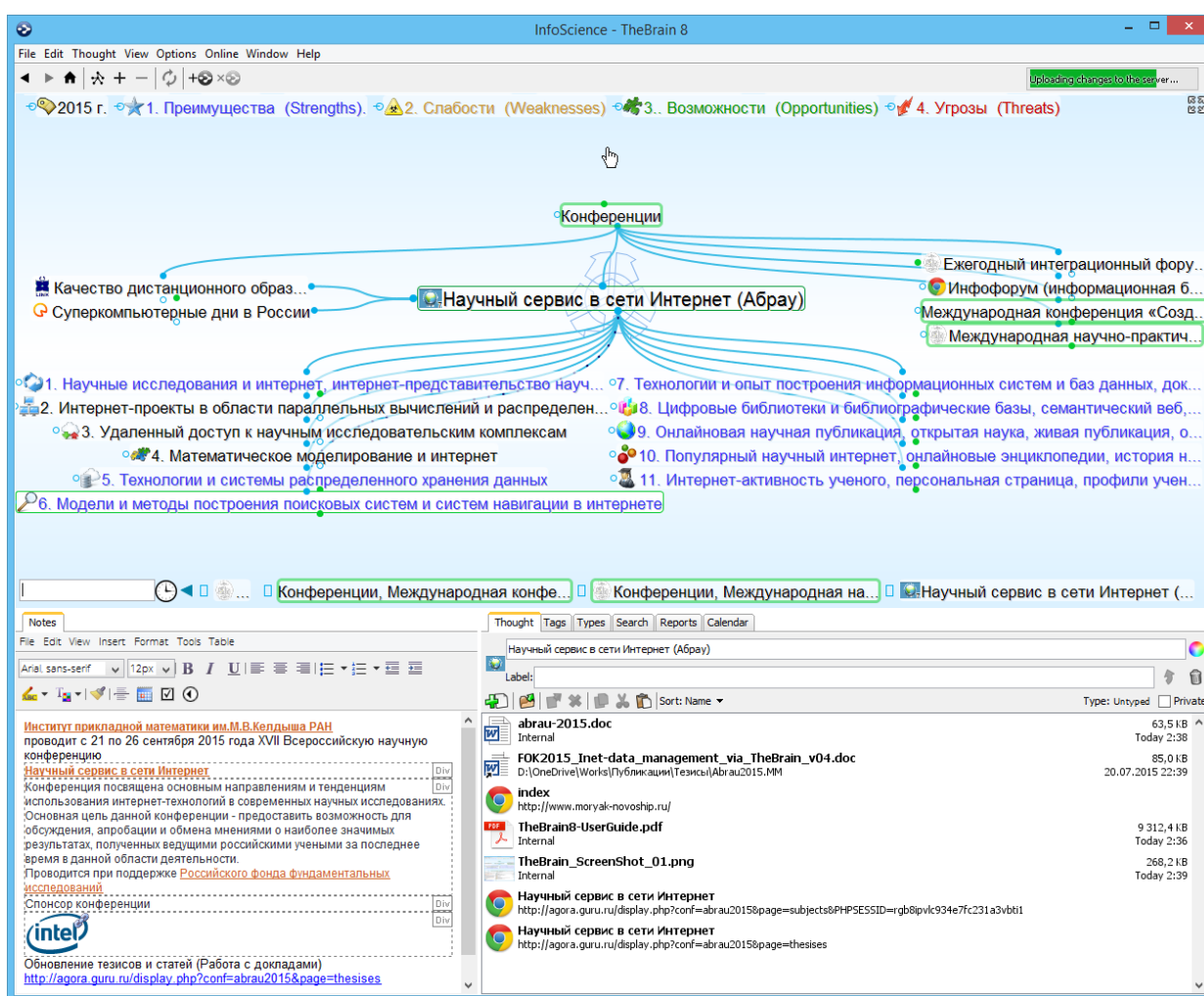


Рис. 1. Экранный снимок графического интерфейса программы TheBrain с ассоциативной картой тематики научных конференций.

Возможность установления связей между элементами разных ассоциативных карт и обращения к ним существенно расширяет свободу управления информационными ресурсами. Интегрированная система экранного

захвата позволяет оперативно ассоциировать изображение произвольной части экрана с элементом карты TheBrain в качестве анонсовой миниатюры или отдельного графического файла. Для максимизации миниатюры до оригинального размера достаточно навести на нее указатель, не вызывая специальную программу просмотра. Веб-страницы с мультимедийным содержимым могут быть не только связаны с элементами посредством гиперссылок, но и вставлены в область заметок элемента, обладающую полноценным инструментарием визуального html-редактора. При необходимости, любой связываемый с элементом файл может быть перенесен или скопирован во внутреннюю структуру ассоциативной карты TheBrain, организованной по принципу распределенных по директориям служебных или подключенных файлов, связанных одной базой данных. Такая структурная организация обеспечивает неограниченное масштабирование ассоциативной карты без существенного снижения скорости доступа к любым элементам и ассоциированным информационным ресурсам. С элементами ассоциативной карты можно связывать почтовые сообщения, контакты и задачи, события и напоминания программы Outlook в среде Windows или программ iCal, Address Book и Apple Mail в среде Mac OS X. Внутренний календарь TheBrain может отслеживать прогресс выполнения связанных с элементами задач и синхронизироваться с Google Calendar, а через него – с календарями платформ Windows и Mac OS.

Важнейшее свойство программы TheBrain – оперативная синхронизация локальных ассоциативных карт с их копиями на защищенном облачном сервере <https://webbrain.com>, предоставляющем 30 Гб для данных одного подписчика. Содержимое этих копий в реальном времени может обмениваться между любым количеством стационарных и мобильных компьютеров на платформах Windows, Apple и Android с соответствующими версиями TheBrain при коллективной работе над одной совокупностью информационных ресурсов. С устройств без установленной программы TheBrain хранящиеся на облачном сервере ассоциативные карты можно просматривать и редактировать из любого браузера. Серверная карта информационных ресурсов может публиковаться по выбору подписчика в открытом или закрытом доступе на любом сайте. В настоящее время разработчики предлагают 4 варианта программы TheBrain с различным функционалом. Наиболее продвинутый вариант с лицензией TeamBrain предполагает возможность сетевого объединения ассоциативных карт разных подписчиков и выделение 100 Гб серверного пространства. Очевидно, что объемные файлы сетевых медиатек целесообразно не включать в структуру ассоциативной карты, а связывать с её элементами гиперссылками на сетевые или локальные копии. К примеру, графические файлы из Flickr и аудиовизуальные ряды из Vimeo целесообразно непосредственно связывать простыми гиперссылками с элементами ассоциативной карты TheBrain, а содержимое неспециализированного облачного хранилища файлов OneDrive с

неутилитарной системой создания ссылок – через локальные синхронизируемые копии.

В течение года мы исследовали функциональные возможности программы ассоциативного картирования TheBrain 8 с лицензией Pro Combo. Неэффективные ассоциативные карты MindView были импортированы в TheBrain через файлы Microsoft Word с иерархически нумерованными списками элементов. Оказалось возможным объединить связанные по смыслу информационные ресурсы, ранее искусственно разобщенные из-за функциональных ограничений MindView. В результате была создана система управления содержимым авторского информационного континуума [1;16;17], состоящего из 21 сайта на 43 доменах, 50 тематических страниц в социальных сетях и 5 облачных ресурсных центров с общей ёмкостью порядка 10 Тб. Эти результаты указывают, что для эффективного управления распределенными информационными ресурсами в настоящее время целесообразно применять программу ассоциативного картирования TheBrain 8 с лицензией Pro Combo или TeamBrain.

В ближайшее время мы планируем использовать TheBrain для объединения в единую базу знаний ранее созданных в MindView пятидесяти ассоциативных карт с материалами научно-исследовательской работы проекта МАСТЕР-МУЛЬТИМЕДИА.

## ЛИТЕРАТУРА

1. Каспаринский Ф.О., Полянская Е.И. Интернет-активность учёного в рамках парадигмы инфоцентризма. // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции (21-26 сентября 2015 г., г. Новороссийск). М.: ИПМ им.М.В.Келдыша. 2015. С. 141-150.
2. Cloud services. // Wikimedia Foundation, Inc. URL: [http://en.wikipedia.org/wiki/Cloud\\_computing#Cloud\\_storage](http://en.wikipedia.org/wiki/Cloud_computing#Cloud_storage) (01.05.2015).
3. Gantt, H.L. Work, Wages and Profit. New York: The Engineering Magazine. 1910. 312 С.
4. Buzan, T. and Buzan, B. The Mind Map book. London: BBC Books. 2000. 320 С.
5. Microsoft Office Project. // Microsoft Corporation URL: <http://office.microsoft.com/ru-ru/project/FX100487771049.aspx> (01.05.2010).
6. Каспаринский Ф.О., Полянская Е.И. Программы интеллектуального картирования как инструмент научной работы. // Научный сервис в сети Интернет: суперкомпьютерные центры и задачи: Труды Международной суперкомпьютерной конференции (20-25 сентября 2010 г., г. Новороссийск). М.: Издательство МГУ. 2010. С. 521-524.
7. ThinkBuzan – Official Mind Mapping software by Tony Buzan. // ThinkBuzan Ltd. URL: <http://www.thinkbuzan.com> (01.05.2010).
8. ConceptDraw MINDMAP: Brainstorming, Mind mapping for Macintosh and Windows. // Computer Systems Odessa. URL: <http://www.conceptdraw.com/en/products/mindmap/> (01.05.2010).

9. MindGenius Mind Mapping Software. // MindGenius Ltd. URL: <http://www.mindgenius.com/> (01.05.2010).
10. MindView – Professional Mind Mapping Software. // MatchWare A/S. URL: <http://www.matchware.com/en/products/mindview/> (01.05.2010).
11. Microsoft Office SharePoint Server. // Корпорация Microsoft. URL: <http://www.microsoft.com/rus/sharepoint/> (01.05.2009).
12. MindManager 8 for Windows – Mind Mapping by Mindjet. // MindJet. URL: <http://www.mindjet.com/products/mindmanager-8-win/overview> (01.05.2010).
13. Mind Map – SimTech MindMapper: Mind Mapping Software for the Visual Mind . // SimTech Systems Inc. URL: <http://www.mindmapper.com> (01.05.2010).
14. TheBrain – Mind Mapping Software, Brainstorming, GTD and Visual KM Software. // TheBrain Technologies LP. URL: <http://www.thebrain.com> (01.05.2010).
15. Лившиц В.М. Скорость переработки информации человеком и факторы сложности среды. // Тарту: ТГУ. 1976. С. 139-146.
16. Каспаринский Ф.О., Полянская Е.И. Многосайтовые системы как инструмент для организации авторского информационного континуума. // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции. М.: Издательство Московского университета. 2012. С. 569-572.
17. Каспаринский Ф.О. Кросспостинг в авторском информационном континууме. // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С. 136-141.

# КРОССПОСТИНГ В АВТОРСКОМ ИНФОРМАЦИОННОМ КОНТИНУУМЕ

Ф.О. Каспаринский<sup>1,2</sup>

*1 Биологический факультет Московского государственного университета*

*имени М.В.Ломоносова*

*2 ООО «МАСТЕР-МУЛЬТИМЕДИА»*

Авторский информационный континуум [1] – информационная среда, объединяющая специализированные интернет-представительства основных видов деятельности организации или физического лица. В соответствии с парадигмой инфоцентризма, для каждой сферы интересов целесообразно формировать собственную информационную среду, структура которой определяется спецификой данных, информации и медиаресурсов [2].

В период экстенсивного развития информационной среды увеличение ареала распространения ресурсов может достигаться посредством размещения сайтов информационного континуума на индивидуальных доменах. Для поддержания устойчивой доступности публикаций в ходе информационного противоборства периода гибридной войны (DDoS-атаки, отключение доменов и пр.) информацию тематических сайтов целесообразно дублировать на нескольких одноименных по второму уровню международных (.com, .edu, .gov, .net, .org, .int, .info, .biz, .name, .pro и др.) и национальных (.ru, .su, .рф) доменах первого уровня (домены-близнецы), которые управляются разными организациями. Подобная политика с 2014 года апробируется в авторском информационном континууме проекта МАСТЕР-МУЛЬТИМЕДИА [1], объединяющем 42 домена, 10 из которых размещают индивидуальные сайты:

1. Медиаресурсы по биологии и медицине: [biomedia.pro](http://biomedia.pro);
  2. Информационные сервисы для биохимиков: [biochemistry.pro](http://biochemistry.pro);
  3. Формирование компетенций по биологии: [biocenter.pro](http://biocenter.pro);
  4. Информационные сервисы для биоэнергетиков: [bioenergetics.pro](http://bioenergetics.pro);
  5. Инфокоонтинуум – среда экспрессии инфоцентристов: [infocontinuum.org](http://infocontinuum.org);
  6. Интернет-представительство проекта МАСТЕР-МУЛЬТИМЕДИА: [master-multimedia.ru](http://master-multimedia.ru);
  7. Медиاماгазин знаний, ресурсов и умений: [mediamagazin.org](http://mediamagazin.org);
  8. Персональный сайт Елены Полянской: [polyanskaya.pro](http://polyanskaya.pro);
  9. Тесты по биологии онлайн: [testbio.pro](http://testbio.pro);
  10. Информационные сервисы для цитологов: [cytology.pro](http://cytology.pro),  
а остальные входят в 11 тематических групп:
1. Интернет-представительство Совета по биологии Учебно-Методического Объединения по классическому университетскому образованию: [bioumo.ru](http://bioumo.ru) + [биоумо.рф](http://bioumo.рф);

2. Теория и практика создания профессиональных видеолекций: [videolecture.pro](http://videolecture.pro) + [videolecture.ru](http://videolecture.ru);
3. Теория и практика видеометода обучения: [videomethod.pro](http://videomethod.pro) + [videomethod.info](http://videomethod.info) + [videomethod.ru](http://videomethod.ru) + [videomethod.pf](http://videomethod.pf);
4. Информист – вестник инфоцентриста: [informyst.ru](http://informyst.ru) + [информист.pf](http://информист.pf);
5. Инфоцентризм – научная концепция и мировоззрение третьего тысячелетия: [infocentrism.com](http://infocentrism.com) + [infocentrism.pro](http://infocentrism.pro) + [infocentrism.ru](http://infocentrism.ru) + [инфоцентризм.pf](http://инфоцентризм.pf);
6. Инфоцентрист – человек третьего тысячелетия: [infocentrist.com](http://infocentrist.com) + [infocentrist.ru](http://infocentrist.ru) + [инфоцентрист.pf](http://инфоцентрист.pf);
7. Персональный сайт Феликса Каспаринского: [kasparinsky.com](http://kasparinsky.com) + [kasparinsky.pro](http://kasparinsky.pro) + [kasparinsky.ru](http://kasparinsky.ru);
8. Медиаколлекция. Онлайн-экспозиция избранных медиаресурсов: [mediacollection.pro](http://mediacollection.pro) + [mediacollection.ru](http://mediacollection.ru) + [медиаколлекция.pf](http://медиаколлекция.pf);
9. Медиакраса информационного пространства: [mediabeauty.pro](http://mediabeauty.pro) + [mediabeauty.ru](http://mediabeauty.ru) + [медиакраса.pf](http://медиакраса.pf);
10. Медиамемориал. Ключи от памяти поколений: [mediamemorial.com](http://mediamemorial.com) + [mediamemorial.ru](http://mediamemorial.ru) + [media-memorial.com](http://media-memorial.com) + [media-memorial.ru](http://media-memorial.ru);
11. Медиаметод передачи знаний – искусство дополняющего распараллеливания информации: [mediamethod.pro](http://mediamethod.pro) + [mediamethod.ru](http://mediamethod.ru) + [медиаметод.pf](http://медиаметод.pf).

Решение о создании доменов-близнецов было инициировано увеличением количества атак на сайты инфоконтинуума, а также блокировкой принадлежавшего инфоконтинууму домена [interes.pro](http://interes.pro). В течение первого квартала 2015 года среднее ежедневное количество посетителей сайтов инфоконтинуума составляло в норме 5000-8000 и превосходило 50000 во время атак при суточном количестве хитов 75000-100000. Анализ показал, что основная масса атак производилась из Америки, Европы и Китая. Инфоконтинуум проекта МАСТЕР-МУЛЬТИМЕДИА создан на основе систем управления содержимым [3] динамических сайтов (Content Management System, CMS) 1С-Битрикс (Управление сайтом, редакция «Бизнес» [4]), которая обеспечивает повышенный уровень информационной безопасности за счет проактивной защиты (контроль целостности, защита сессий, защита редиректов, защита от фреймов, контроль активности, автоматизированный стоп-лист) и встроенного веб-антивируса. Однако вышеперечисленные меры безопасности не помогают против DDoS-атак, вызывающих отказ сервера за счет исчерпания его ресурсов (большинство виртуальных хостингов рассчитано на нагрузку не более 10000 посетителей в сутки). По данным встроенного в CMS инспектора активности сайтов, во время DDoS-атак, которые продолжались от 1 дня до нескольких месяцев, работоспособность отдельных сайтов инфоконтинуума снижалась до 5%. Для противодействия DDoS-атакам домены инфоконтинуума были перемещены с Битрикс-аккаунта на Премиум-аккаунт виртуального хостинга Джино ([jino.ru](http://jino.ru)) для увеличения жесткого лимита нагрузки на CPU сервера с 20% до 30% и размера Memcache с 512 Мб



до 1024 Мб. Был включен композитный режим работы сайтов с кэшированием статических компонентов и активировано ускорение работы сайтов при помощи технологии CDN (географически распределенная сеть доставки и дистрибуции контента). В результате принятых мер работоспособность сайтов поднялась до уровня 80% и сохраняется на этом уровне уже 3 месяца. В ходе противодействия кибератакам было установлена целесообразность использования доменов-близнецов для сохранения доступности информационных материалов: имена интернационализованных доменов отличаются от прочих доменов-близнецов, что снижает эффективность действий информационных атак, направленных на одноименные домены. К примеру, в авторском информационном континууме проекта МАСТЕР-МУЛЬТИМЕДИА информация на тему «инфоцентризм» распространяется между аналогичными сайтами на четырех доменах: infocentrism.ru, infocentrism.com, infocentrism.pro, инфоцентризм.рф (xn--e1aebbvcbgutzs.xn--p1ai) и нормированное ежесуточное количество хитов для этих доменов соотносится как 100:20:2:1.

Мультидоменное тиражирование информации способствует продвижению публикаций вверх в выборке поисковых систем с различными доменными предпочтениями (Search Engine Optimization, SEO). Новые CMS допускают вертикальное масштабирование многосайтового мультидоменного информационного континуума как вниз, так и вверх [1].

Функционал современных CMS позволяет индивидуализировать структуру и дизайн сайтов каждой группы доменов-близнецов в соответствии тематикой, настроить пути распространения авторских публикаций и способы их интеграции с материалами, поступающими от других авторов в информационное пространство специализированных сайтов. Все сайты одной тематической группы должны иметь аналогичную информационную структуру (разделы, страницы, подключенные информационные блоки [5]), а дизайн сайта каждого домена-близнеца допустимо варьировать по мере необходимости. Добавление или редактирование элемента информационных блоков на любом из сайтов-близнецов автоматически приводит к обновлению содержимого всех сайтов данной тематической группы. Имеется возможность организации публикаций аналогичных информационных элементов не только в группе сайтов-близнецов, но и в произвольной совокупности сайтов информационного континуума. К примеру, элементы инфоблоков «Новости проекта МАСТЕР-МУЛЬТИМЕДИА», «Характеристики медиаресурсов», «Авторы информационного континуума» распространяются по всем 42 доменам.

С учетом вероятности усиления кибератак при переходе к жесткой фазе информационного противоборства гибридной мировой войны в 2016 году для обеспечения перманентной доступности информации следует дублировать публикации сайтов интернет-представительств организаций и проектов на специализированных страницах социальных сетей [6]. Традиционный вариант распространения информации с сайтов в социальные сети – комментирование

публикаций посредством штатных или дополнительных инструментов CMS. Однако этот способ неэффективен при большом количестве информационных материалов, поскольку требует отдельных затрат времени на подключение публикации к каждой из социальных сетей. Для автоматизированной передачи информации в социальные сети при помощи свободно распространяемых социальными сетями интерфейсов прикладного программирования (Application Programming Interface, API [7]) создаются специфичные для CMS функциональные модули. Для CMS 1С-Битрикс А.С.Новиковым создан распространяемый на коммерческой основе модуль «Кросспостинг материалов» [8], который позволяет организовать на полностью автоматический кросспостинг (межсайтовую публикацию) любых материалов из инфоблоков сайтов инфоконтинуума сразу в 7 популярных социальных сетей (facebook.com, vk.com, twitter.com, linkedin.com, livejournal.com, blogger.com, ok.com).

Модуль «Кросспостинг материалов» устанавливается штатными средствами CMS 1С-Битрикс. Для корректного функционирования модуля сервер хостинга должен поддерживать работу специальных расширений PHP, а именно: CURL и GD. Добавление и редактирование элементов модуля производится в разделе «Кросспостинг» меню «Сервисы» Административного раздела CMS. В Основных настройках элемента модуля «Кросспостинг» осуществляется подключение инфоблока с элементами, содержимое которых отображается на страницах сайтов инфоконтинуума и предназначено для распространения в социальные сети. Отличительная особенность модуля – добавление в свойства элементов публикуемого инфоблока специальных настроек, контролирующих кросспостинг: разрешение на первичную публикацию в социальной сети, изменение/удаление публикации в соцсети при редактировании/удалении соответствующего элемента инфоблока [9]. Таким образом, контент-менеджер может управлять публикациями социальных сетей из подключенных к ним сайтов инфоконтинуума. При необходимости, может быть включена отсрочка публикации новых или отредактированных элементов инфоблока на определенное время или рандомная публикация всех ранее созданных элементов инфоблока, имеющих разрешение для кросспостинга [8;10].

Для подключения livejournal.com к инфоконтинууму достаточно указать логин и пароль аккаунта социальной сети в соответствующих настройках модуля «Кросспостинг». Остальные социальные сети предъявляют повышенные требования к информационной безопасности и требуют создания специального пользовательского приложения для интеграции с модулем «Кросспостинг». Для каждой из социальных сетей в модуле «Кросспостинг» имеется ссылка для создания соответствующего приложения, управляющего авторизацией в группе доменов-близнецов, административными полномочиями и т.п. В приложении генерируются ключи API, секретные ключи и цифровые подписи (Access Token), которые в настройках модуля дополняют ID API-

приложения и ID страницы (группы) соответствующей социальной сети. Следует контролировать дееспособность цифровых подписей некоторых социальных сетей: в facebook.com и linkedin.com срок годности Access Token ограничен. В модуле «Кросспостинг» имеются инструкции о способах контроля и продления цифровых подписей API-приложений социальных сетей.

Основной функционал модуля «Кросспостинг» позволяет:

- варьировать источник публикуемой текстовой и графической информации для каждой социальной сети (анонс, детальная информация, свойства элемента);
- изменять размер изображений при публикации;
- накладывать водяной знак на публикуемые изображения;
- задавать вид текста публикации индивидуально для каждой социальной сети, используя шаблонизатор [11].

Для каждой социальной сети существует набор специфических настроек публикации, определяющих место публикации (личные или тематические страницы), подпись публикатора (группа/администратор/пользователь), наличие дополнительных изображений, контроль целевой аудитории, формирование URL в публикации социальной сети на исходный материал сайта инфоконтинуума.

Модуль «Кросспостинг» своевременно адаптируется автором к изменениям свойств API социальных сетей. Существенные изменения и дополнения в функциональные возможности модуля были привнесены в ходе экспериментальной интеграции информационного континуума сайтов проекта МАСТЕР-МУЛЬТИМЕДИА с соответствующими 46 страницами в социальных сетях facebook.com, vk.com, twitter.com, linkedin.com, livejournal.com и blogger.com.

В перспективе мы планируем сопоставить эффективность распространения публикаций через сайты инфоконтинуума проекта МАСТЕР-МУЛЬТИМЕДИА и связанные с ними специализированные страницы социальных сетей.

## ЛИТЕРАТУРА

1. Каспаринский Ф.О., Полянская Е.И. Многосайтовые системы как инструмент для организации авторского информационного континуума. // Научный сервис в сети Интернет. Труды Международной суперкомпьютерной конференции. М.: Издательство Московского университета. 2012. С. 569-572.
2. Каспаринский Ф.О., Полянская Е.И. Интернет-активность учёного в рамках парадигмы инфоцентризма. // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С. 142-150.

3. Савельева, Н. Системы управления контентом. // Издательство "Открытые системы". URL: <http://www.osp.ru/os/2004/04/184166/> (01.05.2004).
4. CMS, система управления интернет проектами, создание веб-проектов, система управления внутренним порталом компании – 1С-Битрикс. // СП «1С-Битрикс». URL: <http://www.1c-bitrix.ru> (01.05.2011).
5. Каспаринский Ф.О., Полянская Е.И. Интернет-публикация учебно-методической литературы в форме динамических информационных блоков. // Научный сервис в сети Интернет: Труды Международной суперкомпьютерной конференции. М.: Издательство Московского университета. 2013. С. 504-507.
6. Каспаринский Ф.О., Полянская Е.И. Интернет-представительство научно-образовательных организаций и проектов на специализированных сайтах и в социальных сетях: SWOT-анализ. // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С. 151-159.
7. API // Wikimedia Foundation, Inc. URL: <https://ru.wikipedia.org/wiki/API> (01.05.2015).
8. Новиков, А.С. Кросспостинг материалов. // ООО "1С-Битрикс". URL: <http://marketplace.1c-bitrix.ru/solutions/nsandrey.crossposting/> (01.05.2015).
9. Новиков, А.С. Обновление модуля кросспостинга в соц. сети. // ООО "Дрим Ай Ти". URL: <http://dream-it.su/about-us/news/the-update-module-crossposting-in-sots-network/> (01.05.2015).
10. Новиков, А.С. Версия 2.0 модуля кросспостинга в соц. сети. // ООО "Дрим Ай Ти". URL: <http://dream-it.su/about-us/news/version-2-0-of-the-module-crossposting-in-the-social-network/> (01.04.2015).
11. Новиков, А.С. Примеры использования нового шаблонизатора для кросспостинга. // ООО "Дрим Ай Ти" URL: <http://dream-it.su/articles/examples-of-using-the-new-template-engine-for-crossposting/> (13.05.2015).

# ИНТЕРНЕТ-АКТИВНОСТЬ УЧЁНОГО В РАМКАХ ПАРАДИГМЫ ИНФОЦЕНТРИЗМА

Ф.О. Каспаринский<sup>1,2</sup>, Е.И. Полянская<sup>1</sup>

*1 Биологический факультет Московского государственного университета*

*имени М.В.Ломоносова*

*2 ООО «МАСТЕР-МУЛЬТИМЕДИА»*

Появление сетевых информационных технологий в начале третьего тысячелетия катализировало глобализационную реструктуризацию общественно-экономических и научно-технических процессов. Изменение способов обмена материальными, энергетическими, человеческими и информационными ресурсами инициировало преобразование капитализма в информизм – общественный строй, при котором ключевым ресурсом власти является информация. Фундаментальное изменение свойств жизненной среды человечества вызвало адаптационную модификацию общепринятых и личных парадигм, шаблонизирующих совокупность базовых ценностей, моделей мира или его частей (отраслей, областей знаний, сфер жизни и деятельности), методов принятия решений, подходов, технических навыков и средств [1].

Современная наука пришла к выводу, что живой организм – это машина, управляемая генотипической информацией [2]. Преобладание активно генерируемого исходящего информационного потока отличает живые системы от прочих открытых систем, осуществляющих материальный, энергетический и информационный обмен. Эволюционный отбор генотипических программ осуществляется по фенотипам, обеспечивающим интенсификацию информационного обмена. Эволюционные процессы ускоряются в период кризисов, когда в результате обострения борьбы за существование происходит массовая элиминация неадаптированных организмов и систем. В результате реформирования и последующей модернизации научно-образовательная сфера деятельности была искусственно оптимизирована для коммерческого использования в краткосрочных тактических проектах и замещения творческих исследователей сертифицированными компетентными исполнителями [3-5]. Потребность в свободе научного творчества – результат экспрессии доминирующей генетической программы «учёного по призванию». С наступлением кризиса ряд долгосрочных стратегических фундаментальных исследований утратил поддержку. Можно ожидать, что в течение первых пяти лет после начала Великой экономической депрессии XXI века [6] многим учёным предстоит выбор между добровольной элиминацией из научной среды или активной борьбой за существование в условиях агрессивного перераспределения мировых ресурсов (2015-2020 гг.). Инфоцентрически целенаправленное использование технологий работы с данными, информацией

и медиаресурсами может способствовать своевременной адаптации учёных к глобально модернизирующимся условиям.

## ОПРЕДЕЛЕНИЯ

Парадигма эпохи информизма – инфоцентризм, согласно которому жизнь рассматривается как способ существования, изменения и воспроизводства информации посредством организации частично открытых систем, функционирующих за счет увеличения энтропии окружающей среды в процессе автономного управляемого обмена энергией, материей и данными [1;7]. Это и последующие определения сформулированы на основе 12-летнего опыта экспериментального исследования регулируемого распространения информации в системах биологической энергетики (ведущая организация – Институт проблем передачи информации РАН [8]) и последующей 12-летней научно-практической работы по интеграции информационных и мультимедийных технологий в образовательный процесс [9].

Согласно инфоцентрическому определению, данные – неосмысленные объективные сведения о временном строении и функциях компонентов окружающего мира, перемещаемые носителем в пространстве и времени. При этом под объективностью подразумевается свойство материальных объектов и явлений (процессов) порождать многообразие состояний, которые посредством фундаментальных взаимодействий передаются другим объектам и запечатлеваются в их структуре, обеспечивая потенциальное существование вне и независимо от субъекта. С учетом вышеизложенного, информация – специализированная для передачи в пространстве и времени объективно существующая частично открытая система из прагматически целостных символических описаний структур и функций компонентов Вселенной [1], возникающая за счет энергии и материи окружающей среды в процессе субъективного восприятия, компоновки и интерпретации данных в соответствии с логическим типом (истинная, ложная) и назначением (массовая, специальная, секретная, личная). Медиаресурс – объективный прагматически целостный продукт субъективной адаптации информации для целевой трансляции посредством специализированного распределения между формами представления (текстовая, числовая, графическая, звуковая, динамическая) и сенсорными каналами (визуальный, аудиальный, обонятельный, вкусовой, тактильный, тепловой, болевой, гравитационный, координатный, временной). Прагматическими целями использования информации и медиаресурсов могут быть: обучение, пропаганда, развлечение, переключение внимания и др. Сигнал и сообщение – пространственно-временные совокупности факторов окружающей среды, вызывающие потенциально обратимый переход между состояниями системы, отличающиеся неуправляемостью резкого скачка (сигнал) или возможностью тонкого регулирования (сообщение). В отличие от сигнала, содержимое сообщения может быть непредсказуемым для воспринимающей системы.

Данные, информация, медиаресурсы, сигналы и сообщения передаются через пространство и время совокупностью воспроизводимых состояний носителей (нематериальных и материальных, пассивных и активных, живых и неживых), существующих за счет ресурсов окружающей среды. В отличие от пассивных переносчиков, активные носители способны автономно контролировать стабильность информации и медиаресурсов во времени, адаптируясь к изменениям факторов окружающей среды. Живой носитель (организм) отличается от неживого потенциальной способностью производства оригинальной информации, медиаресурсов и сообщений. Живые носители запрограммированы эволюционировать в направлении увеличения эффективности производства данных, информации и медиаресурсов. Приоритет активности организма определяется наибольшим количеством воспринимаемых сигналов и непредсказуемых сообщений об изменении окружающей среды в результате его действий.

Потребность в исследовательской деятельности – генетическая программа учёного, которая оптимизирует жизненные ценности индивидуума для создания оригинальной информации; поиска и апробации реакции адекватных адресатов новых идей; распространения авторских медиаресурсов. Прогресс при достижении этих целей зависит от решения 7 основных задач: получения, хранения и создания данных, информации и медиаресурсов; публикации оригинальных произведений; аккумуляции реакций на публикацию; формирования авторитетной репутации; финансирования.

## 1. ПОЛУЧЕНИЕ ДАННЫХ, ИНФОРМАЦИИ И МЕДИАРЕСУРСОВ

Эффективность творческой деятельности определяется специфическим инструментарием и условиями его применения. 10-летнее совершенствование веб-сервисов для удаленной работы с данными, информацией и медиаресурсами [4;10] во многих областях науки создало условия для формирования информационного континуума [11], способствующего интеллектуальной деятельности без привязки к определенному рабочему месту. Новый стандарт BYOD (Bring Your Own Device) индуцировал встречную адаптацию к использованию в образовательных, профессиональных и личных целях разноформатных данных и способов их передачи, мобильных HID-устройств и кроссплатформенных приложений, медиарегистраторов и систем дополненной реальности [12]. Оперативное получение, фиксация, систематизация и обобщение необходимых для работы распределенных локальных и сетевых ресурсов стали доступны благодаря появлению утилитарных систем управления знаниями [13].

С другой стороны, стандартизация научных приборов и комплектующих, реактивов и объектов исследования, методов получения и анализа данных, а также способов финансирования научной деятельности воспитала толерантность к ограниченности возможностей, шаблонности действий и поспешности выводов, модификации данных в соответствии с

прагматическими директивами вышестоящих инстанций. Инфоцентрический приоритет учёного – избавление от ограничений и искажений при получении и использовании первичных данных. Современные информационно-коммуникационные технологии могут обеспечивать возможность оперативного получения объективных данных, информации и медиаресурсов из первоисточников, способствовать консолидации единомышленников, сохранению существующих и возникновению новых научных школ.

## 2. ХРАНЕНИЕ ДАННЫХ, ИНФОРМАЦИИ И МЕДИАРЕСУРСОВ

Эволюция цифровых технологий хранения информации в течение последних 30 лет сопровождалась увеличением ёмкости внешних носителей: в 1000 раз при переходе от FD и HD к CD (1982-2003), в 6 раз при смене CD на DVD и Flash (2003-2008), в 5 раз при замене DVD на BD (2008-2013 гг.) и до практической бесконечности при переходе к облачным хранилищам. Недолговечность материальных носителей и устройств для чтения-записи, модернизация аппаратных интерфейсных разъемов, совершенствование стандартов беспроводной передачи данных указывают на целесообразность своевременной репликации версионизируемых копий архивов не только на локальных (оптических, магнитных и твердотельных) носителях с регламентом перезаписи не реже 1 раза в 5 лет, но и в облачных хранилищах с распределением функций хранения и управления. В соответствии с политикой кадровой мобильности, целесообразно организовать синхронизацию данных с личным облачным хранилищем ученого, если это допускается служебным режим секретности. По мнению экспертов в области информационной безопасности, одним из наиболее отказоустойчивых облачных сервисов хранения данных по состоянию на 2015 год является Microsoft OneDrive. В качестве системы управления распределенными локальными и сетевыми ресурсами мы рекомендуем использовать сервис TheBrain [13].

Достоверные данные предшествующих исследований – основа для прогресса фундаментальной науки, эксперименты которой нечасто приносят ожидаемый результат. Аккумуляция достоверных данных происходит в условиях свободного творческого поиска. Потребность грантовой системы финансирования науки в регулярных успешных отчётах катализировала накопление прагматически модифицированных данных, препятствующих продуктивной работе последователей. Экстраполяция графиков волновых общественно-экономических и научно-технических процессов на ближайшее будущее позволяет предсказать, что в период с 2021 по 2027 год количество знаний возрастет в 2.5 раза благодаря системному кризису: избавленные от волюнтаристских директив творцы-теоретики работают продуктивнее [1]. Можно ожидать, что организация сетевых профессиональных эпистемотек поможет экспертному научному сообществу избавиться от недостоверных данных, накопившихся в информационном пространстве за 30 лет функционирования грантовой системы.



### 3. СОЗДАНИЕ АВТОРСКИХ ПРОИЗВЕДЕНИЙ (ИНФОРМАЦИИ И МЕДИАРЕСУРСОВ)

Передавая опыт исследовательской работы студентам Биологического факультета МГУ имени М.В.Ломоносова, основоположник отечественной школы биохимиков академик С.Е.Северин говорил: «Гениальные мысли приходят ко всем, но пользу они приносят только тем, кто оказался дома и своевременно открыл дверь». Данные – кирпичи дома интересных идей, в котором автор создает новую информацию и адаптирует её для внешнего мира, облекая в форму медиаресурсов. Идея, своевременно не подкрепленная подходящими данными, уходит из информационного пространства. Для мозгового штурма значительно эффективнее использовать современные средства управления распределенными данными, информацией и медиаресурсами [11;13], нежели блокнот и карандаш. Для каждой сферы научных интересов целесообразно формировать собственную информационную среду, структура которой определяется спецификой данных, информации и медиаресурсов. При необходимости междисциплинарной интеграции специализированные ресурсные центры могут быть объединены в авторский информационный континуум [11;14].

Традиционные инструменты для коллективной работы с данными в современной сетевой среде дополнились возможностями систем поиска мультимедийных материалов с возможностями контролирования прав доступа к результатам (Facebook Graph Search, Google Search и др.). Многие технологии для работы с данными, информацией и медиаресурсами уже адаптировались для использования в мобильных устройствах, обеспечив переход на качественно новый уровень авторского комфорта при создании произведений.

Известно, что экономическая депрессия катализирует увеличение количества знаний и появление новых технологий [1]. Можно ожидать, что конвергенция nano-, био-, информационных и когнитивных технологий (NBIC) в период с 2021 по 2027 гг. поможет преодолеть системный кризис, возродит исследовательский интерес к материальным технологиям и кибернетике. Созданные в годы кризиса теоретические результаты могут востребоваться учеными-практиками в период экономического подъема (2028-2033 гг.) и участвовать в формировании инфоцентрического самосознания нового поколения.

### 4. ПУБЛИКАЦИЯ ОРИГИНАЛЬНЫХ ПРОИЗВЕДЕНИЙ

Первоначальная цель авторской публикации – адаптация оригинальной информации к максимально широкому распространению в пространстве и времени. Традиционные способы издания авторских произведений посредством материальных носителей (книги, журналы, кассеты, оптические диски и т.д.) уходят в прошлое, поскольку ограничиваются юридическими нормами тиражирования их экземпляров [15]. Современные издательства в дополнение к

материальным публикациям начали создавать цифровые варианты произведений и их частей в форматах, приспособленных для сети или мобильных приложений. Присвоение Идентификатора Цифрового Объекта (DOI) обеспечивает достоверность и актуальность библиографических данных в лабильной сетевой среде, однако юридически препятствует потенциальной модификации публикаций авторами [18]. Благодаря эволюции систем управления веб-контентом (CMS) [16] у авторов появилась возможность распространять подконтрольные им живые публикации [17] через порталы организаций, собственные веб-сайты или специально адаптированные социальные сети (Facebook, LinkedIn и др.). Для удостоверения авторства живых публикаций на порталах организаций или персональных авторских сайтах могут использоваться специальные сервисы, позитивно влияющие на позицию оригинальной публикации в выдаче поисковых служб [19].

Результаты мониторинга агрессивных действий компьютерных злоумышленников, нарушающих работу сетевых информационных систем, указывают на целесообразность распространения копий важных публикаций между несколькими сервисами с высокой устойчивостью к DDoS-атакам. К примеру, весной 2015 года ежедневное количество посетителей 40 сайтов информационного континуума проекта «МАСТЕР-МУЛЬТИМЕДИА» [14] превышало 50000, тогда как системные ресурсы большинства виртуальных хостингов рассчитаны на нагрузку не более 10000 посетителей в сутки.

В течение последних двух лет существенно возросла активность публикаторов-временщиков, извлекающих выгоду из искусственно культивируемой потребности учёных опубликовать свои труды в рецензируемых изданиях с высоким импакт-фактором, которые регистрируются в базах данных Web of science, Scopus, Thomson Reuters и РИНЦ. Наличие специализированных сайтов издателей и связанных с ними страниц в социальных сетях с соответствующими постоянными адресами может указывать на благонадежность публикаторов.

## 5. АККУМУЛЯЦИЯ РЕАКЦИЙ НА ПУБЛИКАЦИЮ

За 10 лет существования веб-сервисов сформировался набор сигналов (статусы «Я рекомендую», «Мне нравится», «лайки» и т.п.) и вариантов сообщений (рейтинги, опросы, голосования, отзывы, комментарии и т.п.), информирующих автора и общественность о реакции на публикацию. Приложения для мобильных устройств ускоряют обмен мнениями и организацию ответных действий (мероприятия, организация сообществ, подготовка коллективных трудов и пр.).

Современные средства анализа эффективности сетевой публикации позволяют совершенствовать способы распространения информации посредством вариаций типов публикаций, целевой аудитории (география, язык, социальный статус, группировки по интересам и т.д.).

Аккумуляция читательских реакций способствует повышению позиций публикации в поисковой выдаче при использовании поисковых систем для сбора данных, создает условия для повышения индекса цитирования и препятствует отключению тематических страниц социальных сетей вследствие отсутствия активности, критическая продолжительность которого определяется пользовательским соглашением.

## 6. СОЗДАНИЕ АВТОРИТЕТНОЙ РЕПУТАЦИИ

Помимо содержимого авторских произведений, на репутацию и карьеру учёного в 2015 году начали влиять персональные рейтинги, рассчитываемые на основании наукометрических данных, своевременно вводимых сотрудником в специальные веб-формы сайта организации-работодателя. Формулы для определения рейтингов специфичны для структурных подразделений организации и могут содержать показатели результативности научной работы сотрудника (публикации, монографии, участие в конференциях и пр.), учебной работы (чтение лекций, научное руководство и пр.), привлечённого финансирования (гранты, договора и пр.) за разные периоды времени (от одного года до пяти лет). Определение формулы (показатели, их вес и соотношения) и последующий расчёт рейтингов является инструментом, позволяющим провести оценку степени соответствия сотрудников квалификационным критериям при первичном и повторном конкурсном избрании на научные должности. По замыслу создателей рейтинговой системы, значение персонального рейтинга научного сотрудника не является мерой признания конкретного учёного и не может использоваться для его сравнения с коллегами. Рейтинговые индексы ученого в наукометрическом сервисе могут расти по мере увеличения числа зарегистрированных и коммерчески реализованных публикаций, а также количества перекрестных ссылок между ними. В настоящее время полезно наблюдать за ходом конкуренции между сервисами, присваивающими учёным идентификационные номера (ResearcherID, ORCID, IRID и пр.).

Интернет-активность ученого может влиять на его репутацию: при выборе адресов электронной почты и доменов для размещения сайтов авторских публикаций следует избегать бесплатных сервисов. В научных школах старой формации реакция на корреспонденцию может зависеть от соблюдения негласных правил деловой переписки, таких как обращение «Глубокоуважаемый!» к сотруднику МГУ имени М.В.Ломоносова.

Содержимое портфолио учёного с хронологией ключевых событий и списками научных трудов целесообразно публиковать на служебном и персональном сайтах, а также в сетевых энциклопедиях (wikipedia.org) и авторитетных социальных сетях (facebook.com). В период подготовки к точке технологической сингулярности (2034-2039 гг.) своевременное формирование портфолио может оказаться судьбоносным фактором.

## 7. ФИНАНСИРОВАНИЕ ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ

Живая система действует благодаря самоорганизации автономно управляемого обмена энергией, материей и данными с внешней средой. В соответствии с инфоцентрическими представлениями, деньги – одна из форм свободной энергии. Известно, что длительное ограничение энергоснабжения живой системы может активировать программу ее самоуничтожения. В условиях развития общемирового системного кризиса на фоне смены мировоззренческих парадигм и эскалации мировой гибридной войны целесообразно использовать коммерческий потенциал современных веб-сервисов для конъюнктурно-независимого финансирования фундаментальных научных исследований и живых публикаций. Для получения оплаты за выполнение научных и опытно-конструкторских работ, профессиональные консультации и выступления, создание и распространение медиаресурсов, а также приёма благотворительных пожертвований целесообразно учреждение Обществ с ограниченной ответственностью (ООО) и последующее заключение договоров на получение наличных и безналичных платежей через интернет-сервис «Яндекс.Касса». При планировании коммерческой деятельности следует учитывать, что исключительные права на служебное произведение возвращаются к автору от работодателя, который в течение 3 лет с момента создания произведения не начал его использование и официально не запретил публикацию (Гражданский Кодекс РФ, ст. 1270, ч.2 и ст.1295, ч.2).

## ВЫВОД

Таким образом, все виды деятельности современного учёного связаны с интернет-активностью, которая приобретает роль ключевого фактора адаптации к условиям информационного общества.

## ЛИТЕРАТУРА

1. Каспаринский, Ф.О., Полянская, Е.И. Инфоцентризм как дидактическая стратегия // Вестник Международного ин-та менеджмента ЛИНК (5). М.: МИМ ЛИНК. 2014. С. 65-73.
2. Dawkins,R. The selfish gene. Oxford: Oxford University Publishing, 1976. 360 С.
3. Blurton, C. New Directions of ICT-Use in Education. URL: <http://www.unesco.org/education/educprog/lwf/dl/edict.pdf> (06.02.2007).
4. Каспаринский Ф.О., Полянская Е.И. Дидактически целенаправленное использование информационного инструментария // Качество дистанционного образования: концепции, проблемы, решения. М.: МГИУ, 2012. С. 74-85.
5. Каспаринский, Ф.О., Полянская, Е.И. Адаптация ресурсов дистанционного обучения к компетентностному формату // Открытое образование (4). М.: МЭСИ. 2014. С. 11-19.

6. Каспаринский Ф.О., Полянская Е.И. Прогноз проблем дидактики на основе взаимосвязи экономических волн Кондратьева и смены поколений // Качество дистанционного образования: концепции, проблемы, решения. М.: МГИУ. 2013. С. 78-82.
7. Skulachev, V.P., Bogachev, A.V., and Kasparinsky, F.O. Principles of Bioenergetics. Heidelberg: Springer. 2013. 470 С.
8. Каспаринский, Ф.О. Кинетика электрогенного транспорта двухвалентных катионов в митохондриях. Автореферат диссертации на соискание ученой степени кандидата биологических наук. Ведущая организация: Институт Проблем передачи информации РАН. 2000. 22 С.
9. Каспаринский Феликс Освальдович – пользователь, сотрудник | ИСТИНА - Интеллектуальная Система Тематического Исследования НАучно-технической информации. // Лаборатория 404 НИИ механики МГУ URL: <http://istina.msu.ru/profile/Kasparinsky/> (01.05.2015).
10. Cloud services // Wikimedia Foundation, Inc. URL: [http://en.wikipedia.org/wiki/Cloud\\_computing#Cloud\\_storage](http://en.wikipedia.org/wiki/Cloud_computing#Cloud_storage) (01.05.2015).
11. Каспаринский Ф.О., Полянская Е.И. Многосайтовые системы как инструмент для организации авторского информационного континуума // Научный сервис в сети Интернет. Труды Международной конф. М.: Изд-во МГУ. 2012. С. 569-572.
12. Дендев, Б. Информационные и коммуникационные технологии в образовании. М.: ИИТО ЮНЕСКО. 2013. 320 С.
13. Каспаринский Ф.О. Использование программ ассоциативного картирования для управления распределенными информационными ресурсами // Научный сервис в сети Интернет: Труды XVII Всеросс. науч. конф. М.: ИПМ им.М.В.Келдыша. 2015. С. 128-135.
14. Каспаринский Ф.О. Кросспостинг в авторском информационном континууме // Научный сервис в сети Интернет: Труды XVII Всеросс. науч. конф. М.: ИПМ им.М.В.Келдыша. 2015. С. 136-141.
15. Федеральный закон от 29.12.1994 № 77 "Об обязательном экземпляре документов" (с изменениями на 26 марта 2008 года). URL: [http://www1.fips.ru/wps/wcm/connect/content\\_ru/ru/documents/russian\\_laws/federal\\_statute\\_rf/fed\\_zak\\_77\\_29121994](http://www1.fips.ru/wps/wcm/connect/content_ru/ru/documents/russian_laws/federal_statute_rf/fed_zak_77_29121994) (01.05.2008).
16. Савельева, Н. Системы управления контентом // Издательство "Открытые системы" URL: <http://www.osp.ru/os/2004/04/184166/> (01.05.2004).
17. Полилова Т.А. Библиографическая ссылка и живая публикация // Научный сервис в сети Интернет: Труды Международной конф. М.: Изд-во МГУ, 2010. С. 448-452.
18. Paskin, N. Digital Object Identifier (DOI®) System // Taylor & Francis URL: <http://www.doi.org/overview/080625DOI-ELIS-Paskin.pdf> (01.05.2010).
19. Оригинальные тексты // ООО "Яндекс" URL: <https://webmaster.yandex.ru/content/?service=ORIGINALS> (01.05.2015).

# ИНТЕРНЕТ-ПРЕДСТАВИТЕЛЬСТВА НАУЧНО-ОБРАЗОВАТЕЛЬНЫХ ОРГАНИЗАЦИЙ И ПРОЕКТОВ НА СПЕЦИАЛИЗИРОВАННЫХ САЙТАХ И В СОЦИАЛЬНЫХ СЕТЯХ: SWOT-АНАЛИЗ

Ф.О. Каспаринский<sup>1,2</sup>, Е.И. Полянская<sup>1</sup>

*1 Биологический факультет Московского государственного университета  
имени М.В.Ломоносова*

*2 ООО «МАСТЕР-МУЛЬТИМЕДИА»*

С 2000 года представительство научно-образовательных организаций в сети Интернет стало нормой, а с 2008 года – обязанностью. В соответствии с Приказом №1770 от 28.11.2008 г. Министерства Образования и Науки РФ, регулярное обновление сайта организации (не реже 2 раз в месяц для каждого подразделения) было включено в показатели результативности научно-образовательной деятельности. В результате, научно-образовательные организации оказались вовлечены в процесс естественного отбора технологий, обеспечивающих расширение ареала информационного присутствия в сети Интернет.

Первые представительства научно-образовательных сайтов в Сети представляли собой свёрстанные вручную одностраничные статические сайты-визитки, которые посредством горизонтального масштабирования разрастались в совокупность связанных гиперссылками html-страниц, не адаптированных к обновлению сведений и вариабельной демонстрации. В процессе развития официального сайта Биологического факультета МГУ имени М.В.Ломоносова с 2000 по 2004 гг. обнаружилось, что посредством ручной верстки удаётся эффективно поддерживать сайт не более чем из 300 страниц. Использование конструкторов статических сайтов с поддержкой множественных мастер-страниц увеличивает предел количества эффективно администрируемых страниц до 600 шт. Поскольку в структуре Биофака МГУ находится более 30 подразделений, ратификация нормативов 2008 года сокращает продолжительность жизненного цикла статического сайта до одного года. Именно поэтому с 2010 г. большинство научно-образовательных сайтов поддерживается при помощи более производительных технологий динамических облачных публикаций [1], осуществляемых Системами Управления Содержимым (Content Management Systems, CMS) [2;3]. С 2004 года CMS стали использоваться для создания и администрирования социальных сетей (Social Media), приспособленных для организации сообществ индивидуальных публикаторов медиаресурсов и аккумуляции реакций на публикации [4].

Эволюционная специализация социальных сетей способствовала расширению областей их применения от развлекательных до рекрутинговых, образовательных и научных. Средства массовой информации (теле-радиоканалы, издательства и пр.) и ИТ-компании первыми стали распараллеливать информацию о себе между своими официальными интернет-сайтами и страницами (сообществами, группами) в социальных сетях. МГУ имени М.В.Ломоносова (msu.ru) открыл свои официальные страницы в facebook.com (с 24.09.2012 <https://www.facebook.com/MoscowStateUniversity>), twitter.com (с 14.08.2013 [https://twitter.com/VA\\_Sadovnichiy](https://twitter.com/VA_Sadovnichiy)) и vk.com (с 02.10.2014 [https://vk.com/msu\\_official](https://vk.com/msu_official)).

Многие научно-образовательные организации продолжают откладывать решение об открытии интернет-представительств в социальных сетях, обобщённо представляемых в общественном мнении как облачные сервисы для развлекательного самовыражения безответственных анонимов. Безусловно, анонимность информационного обмена контрпродуктивна в зависимой от авторитета экспертов научно-образовательной сфере. Для преодоления проблемы анонимности социальная сеть Facebook.com с 30 января 2015 года ввела жесткие меры: первичная регистрация невозможна без проверки номера мобильного телефона пользователя, а доступ к аккаунту в любой момент может быть заблокирован в результате отказа от такой проверки, которая инициируется при наличии подозрительной активности. Ведущие социальные сети (twitter.com, vk.com) вскоре ввели у себя подобный механизм идентификации пользователей. Таким образом, настало время ревизии целесообразности использования социальных сетей в качестве информационного представительства научно-образовательных организаций и проектов в сети Интернет.

В этой статье сопоставляются варианты интернет-активности на официальных сайтах научно-образовательных организаций и на страницах социальных сетей посредством SWOT-анализа — метода стратегического планирования [5], который сегрегирует факторы внутренней и внешней среды организации (виды деятельности) на четыре категории: Strengths (сильные особенности), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы). Сильные (S) и слабые (W) стороны являются подконтрольными объекту анализа внутренними факторами, а возможности (O) и угрозы (T) представляют собой факторы внешней среды, потенциально влияющие на объект. Принятие стратегического решения зависит от результатов распределения факторов по областям:

- SO показывает, какие сильные внутренние стороны необходимо использовать, чтобы получить отдачу от возможностей во внешней среде;
- WO аккумулирует возможности внешней среды, способствующие преодолению внутренних слабостей;
- ST сосредоточивает силы, которые необходимо использовать для устранения угроз;

- WT концентрирует слабости, от которых необходимо избавиться, чтобы попытаться предотвратить актуальную угрозу.

## STRENGTHS (ВНУТРЕННИЕ СИЛЬНЫЕ ОСОБЕННОСТИ)

Практика проведения аккредитационных экспертиз в течение последних 5 лет указывает, что основная оценка конкурентоспособных преимуществ организации или проекта производится на основании анализа содержимого общедоступного интернет-представительства. По этой причине работа по своевременному созданию, сохранению и распространению благоприятного образа организации (проекта) в информационном пространстве (брендинг) является необходимым условием для выживания в условиях нынешнего системного кризиса. Формирование бренда организации включает пропаганду официальной символики (герб, флаг, гимн, девиз, логотип), наполнение портфолио достижений, публикацию нормативных документов и сертификатов, размещение персональных страниц сотрудников, составление исторических хроник, а также романтизацию неофициальных традиций. Современные CMS-конструкторы сайтов и передовые социальные сети предоставляют все возможности для эффективного брендинга, включая индивидуализацию дизайна. С прошлого года ссылки на страницы официального сайта организации (проекта) в статьях социальных сетей рассматриваются как серьезный фактор оптимизации репутационного продвижения информации в поисковых системах (Search Engine Optimization, SEO). При создании страниц в социальных сетях следует адаптировать графическую атрибутику организации (проекта) к индивидуальной специфике сети вследствие отсутствия единого стандарта. Существование конкуренции за имена страниц является фактором, стимулирующим своевременную организацию страниц организаций (проектов) в социальных сетях и их последующую популяризацию. К примеру, произвольное присвоение имени страницы в социальной сети Facebook.com становится доступно только после получения 25 одобрительных сигналов «Нравится» [4] от персональных страниц индивидуальных пользователей, причем сигналы от тематических страниц не принимаются в расчёт.

В соответствии со структурно-функциональной спецификой научно-образовательной организации (проекта) для каждого подразделения целесообразно создавать индивидуальный раздел сайта с собственным информационным блоком или специальный сайт, публикации которых следует распараллеливать на соответствующие специальные страницы каждой из ведущих социальных сетей (facebook.com, vk.com, twitter.com, linkedin.com, blogger.com, livejournal.com, plus.google.com и др.). Тематические страницы социальных сетей посредством вертикального масштабирования вниз можно дополнять страницами научных проектов и учебных структурных подразделений, а также авторскими страницами с портфолио сотрудников или учащихся. Альтернативный способ кадровой консолидации – подключение личных страниц к дочерним группам, созданным из страницы подразделения



организации. Такая структуризация облегчает доступ к материалам в соответствии с их тематикой, поскольку одновременная тематическая фильтрация информации разных типов (текст, графика, анимация, аудио, видео) из общей ленты публикаций в социальных сетях может отсутствовать или функционировать неэффективно. Автоматическое центробежное распространение публикаций от официального сайта организации (проекта) в социальные сети может обеспечиваться специфичными для CMS сайта средствами кросспостинга [6], создаваемыми на основе общедоступных интерфейсов прикладного программирования (API) социальных сетей [7]. Правильная организация кросспостинга позволяет при редактировании публикации центрального сайта не дублировать и не уничтожать ранее опубликованные материалы, а модифицировать их. Перекрестные публикации между социальными сетями также возможны, но они приводят к контрпродуктивному дублированию публикаций при их редактировании.

Современные CMS сайтов допускают любые типы масштабирования, включая горизонтальное и конвергентное масштабирование вверх. Функционал социальных сетей допускает вертикальное масштабирование вниз, если тематические страницы создаются одним администратором, а также горизонтальное масштабирование пропорционально количеству администраторов с индивидуальными аккаунтами, подтвержденными разными номерами мобильных телефонов. Некоторые социальные сети лимитируют количество страниц, создаваемых одним администратором за один день (не более 20 в vk.com) или за всё время существования аккаунта (не более 10 в linkedin.com). Большинство социальных сетей допускает структурно-функциональную адаптацию дизайна тематических страниц к их специфике.

Инструментарий для публикации информации различных типов уже достаточно совершенен как у индивидуальных CMS-поддерживаемых сайтов, так и у социальных сетей. Как правило, публикации поддерживают мультимедийность (прикрепление фото, видео, аудио, комментариев) и интерактивность (аккумуляция пользовательских сигналов и сообщений) [4]. Статистический анализ эффективности сетевых публикаций на страницах бизнес-версий CMS-сайтов и в социальных сетях позволяет хронологизировать востребованность различных типов материалов, заинтересованность целевой аудитории (география, язык, социальный статус, группировки по интересам и т.д.), пользовательскую реакцию (сигналы и сообщения) и пр. Поиск текстовой информации в социальных сетях может быть дополнен системой распознавания мультимедийных материалов с возможностями контролирования прав доступа к результатам (Facebook Graph Search, Google Search и др.). Социальная сеть facebook.com предлагает сервис «Взгляд в прошлое», посредством которого анализируются все новости страницы и автоматически компонуется рекламный фильм, подборка фотографий или открытка с благодарностью. При необходимости, созданные сервисом медиаресурсы могут быть оперативно отредактированы и использованы для брендинга и включения в отчёты.

Современные технологии обеспечивают передачу информационных материалов с дополнительными данными (названия, описания, комментарии, геопривязка, метки лиц) не только посредством традиционных веб-форм, но и непосредственно из приложений мобильных устройств. В настоящее время эволюционирует функционал структурирования информационного содержимого (контролируемое пользователями создание древовидных структур разделов, альбомов, тематических выборок и т.п.). Перцептивно-комфортное отображение медиаматериалов на экранах стационарных и мобильных устройств обеспечивает адаптивный дизайн, принципы которого совершенствуются уже 5 лет [8].

Распространение публикаций на страницах официальных сайтов и в социальных сетях можно контролировать с учетом местонахождения корреспондента, его принадлежности к определенной профессиональной, возрастной, социальной группе и пр. Фильтрация корреспондентов в социальных сетях по IP адресам пока не применяется, что формально препятствует публикации на их страницах конфиденциальной и коммерчески значимой информации. Возможности информационного обмена через публикации и их приложения на специализированных сайтах и в лидирующих социальных сетях дополняются наличием менеджера сообщений (чат), который позволяет в реальном времени индивидуально или циркулярно передавать текст, изображения, звук и видео, файлы в любых форматах и организовывать видеоконференции, не используя сторонние сервисы. Таким образом, комфортные условия для коллективной работы могут быть созданы как на официальных сайтах организаций, так и в социальных сетях.

## WEAKNESSES (ВНУТРЕННИЕ СЛАБЫЕ СТОРОНЫ)

В отличие от сайтов, созданных при помощи CMS бизнес-класса, в социальных сетях пока нет функционала для создания баз знаний (wiki), организации сложных опросов и голосований, сбора наукометрической информации и первичных данных, управления бизнес-процессами и документооборота. Основные социальные сети адаптированы для информирования подписчиков о публикациях через приложения мобильных устройств, тогда как для организации подобного сервиса на официальном сайте организации (проекта) потребуется специальный модуль для CMS и подписка на внешний сервис информационного обмена. Социальные сети позволяют эффективно организовывать мероприятия, фиксировать ключевые события и оперативно информировать их участников о новостях.

При редактировании элементов информационных блоков сайтов с поддержкой бизнес-процессов возможно организовать версионирование документов, примитивный аналог которого уже работает при редактировании публикаций в социальных сетях.

## OPPORTUNITIES (ВНЕШНИЕ ВОЗМОЖНОСТИ)

Ведущие социальные сети обладают функционалом поиска вероятных коллег, партнеров и знакомых, который способствует решению вопросов трудоустройства подготовленных в организации специалистов, привлечения новых кадров, создания партнерских программ, получения профессиональных консультаций, рекламы произведений и услуг, поиска источников финансирования и административных ресурсов. При необходимости, социальные сети становятся важным источником контактной информации правопреемников обладателей авторских, исключительных и смежных прав на продукты интеллектуальной деятельности.

Социальные сети на платной основе могут стимулировать рост общественного интереса к организации (проекту), укрепляя его репутацию и расширяя охват целевой аудитории. Интернет-представительства способствуют получению дополнительного финансирования организаций и проектов от коммерческой реализации произведений и услуг, размещения рекламы и приема благотворительных пожертвований. Для этого требуется разрешение на осуществление соответствующих видов деятельности согласно ОКВЭД и открытие интернет-магазинов на официальном сайте или в социальных сетях. Перевод безналичных денежных средств с пластиковых карт или из электронных кошельков Яндекс.Деньги на расчетный счёт организации можно реализовать посредством размещения конструируемой формы сервиса быстрых платежей Яндекс.Деньги (<https://money.yandex.ru/fastpay>) на странице сайта или социальной сети.

## THREATS (ВНЕШНИЕ УГРОЗЫ)

Оперативность публикаций на официальном сайте организации (проекта) зависит от загруженности его администратора, несущего ответственность за распространение информации. Размещение материалов в социальных сетях и их редактирование происходит в реальном времени под контролем их автора. При необходимости, автор может разрешить коллективное редактирование опубликованных им материалов. Если инструменты модерирования публикации и установления корреспондентов отсутствуют или деактивированы, нежелательная или конфиденциальная информация может оказаться временно доступной для свободного ознакомления. Социальные сети не всегда предоставляют возможность модерирования комментариев и отзывов на публикации, что может нести угрозу распространения балластной, чуждой и вредоносной информации. Реклама в социальных сетях может служить дистрактором внимания, ухудшающим восприятие публикаций.

Развитие технологий требует своевременного обновления CMS официального сайта, что не всегда оказывается возможным в условиях госбюджетного финансирования. Жесткая конкуренция между социальными

сетями вынуждает их предоставлять самые современные сервисы и своевременно искоренять выявленные недостатки.

Актуальная проблема – обеспечение доступности сетевых публикаций. Для преодоления последствий вероятного отключения отдельных доменных зон в результате введения политико-экономических санкций целесообразно дублировать публикации официальных сайтов организаций (проектов) на одноименных доменах различных зон и во множестве социальных сетей. Подобное клонирование публикаций может также рассматриваться как способ поддержания доступности информации во время направленных на официальные сайты организаций DDoS-атак, частота и сила которых увеличивается в соответствии со сценарием ведения гибридной мировой войны за перераспределение ресурсов. Существует вероятность, что в 2016 году обеспечение информационной безопасности станет главной заботой администраторов интернет-представительств.

Смена владельца, домена, хостинга, администратора, CMS и структуры официального сайта может сопровождаться утратой ранее существовавших страниц, как это произошло с официальными сайтами Биофака МГУ (в 2004 и 2009 годах) и Московского государственного университета имени М.В.Ломоносова (в 2014 году). К примеру, реорганизация сайта Биологического факультета МГУ привела к утрате персональных страниц декана факультета (1973-2005) профессора М.В.Гусева (1935-2005), заведующего кафедрой Физиологии человека и животных (1986-2006) академика РАМН И.П.Ашмарина (1925-2007), заведующего кафедрой Высших растений (1972-1977) профессора А.П.Меликяна и ряда других. Аналогичные проблемы могут возникать в результате волюнтаристских административных решений о ликвидации независимых сайтов подконтрольных организаций с целью подконтрольного слияния их информации на сайте головной организации. Оптимальный вариант для сохранения ценной информации – её публикация в социальной сети, лицензионное соглашение с которой не предусматривает уничтожение аккаунта при отсутствии активности его администратора в течение определенного периода времени. Однако при эволюции социальных сетей может происходить временное отключение или уничтожение сервисов, что может потребовать перемещения публикаций на новое место.

## ВЫВОДЫ И РЕКОМЕНДАЦИИ

1. Для получения отдачи от возможностей во внешней среде целесообразно участвовать в первичном распределении доступных оптимальных имён страниц социальных сетей и организовать центробежный кросспостинг от официального сайта к специализированным страницам социальных сетей для повышения рейтинга организации (проекта).
2. Функциональные возможности социальных сетей полезно использовать для улучшения оперативности распространения информации среди сотрудников

организации (проекта), профессиональной консолидации, привлечения внимания инвесторов и административных ресурсов.

3. Регулярный анализ эффективности публикаций на официальных сайтах и в социальных сетях рекомендуется для определения тенденций изменения отношения к организации (проекту) и их своевременной корректировки.
4. Для обеспечения информационной безопасности необходимо оперативно реагировать на соответствующие рекомендации администраторов интернет-представительств и активно пользоваться настройкой конфиденциальности в процессе публикации.

SWOT-анализ – эффективное средство для определения стратегии развития интернет-представительств научно-образовательных организаций и проектов. В качестве инструмента для осуществления этой работы удобно использовать программы ассоциативного картирования [9]. Целесообразно повторение и углубление SWOT-анализа использования интернет-представительств в соответствии с актуальными изменениями свойств научно-образовательной и информационной сфер деятельности.

## ЛИТЕРАТУРА

1. Cloud services // Wikimedia Foundation, Inc. URL: [http://en.wikipedia.org/wiki/Cloud\\_computing#Cloud\\_storage](http://en.wikipedia.org/wiki/Cloud_computing#Cloud_storage) (01.05.2015).
2. Савельева, Н. Системы управления контентом // Издательство "Открытые системы". URL: <http://www.osp.ru/os/2004/04/184166/> (01.05.2004).
3. Каспаринский Ф.О., Полянская Е.И. Интернет-публикация учебно-методической литературы в форме динамических информационных блоков // Научный сервис в сети Интернет: Труды Международной суперкомпьютерной конференции. М.: Издательство Московского университета. 2013. С. 504-507.
4. Каспаринский Ф.О., Полянская Е.И. Интернет-активность учёного в рамках парадигмы инфоцентризма // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С. 141-150.
5. Майсак,О.С. SWOT-анализ: объект, факторы, стратегии. Проблема поиска связей между факторами. Прикаспийский журнал: управление и высокие технологии. 2013. т.1. С. 151-157.
6. Каспаринский Ф.О. Кросспостинг в авторском информационном континууме // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С. 136-141.
7. API // Wikimedia Foundation, Inc. URL: <https://ru.wikipedia.org/wiki/API> (01.05.2015).
8. Каспаринский Ф.О., Полянская Е.И. Дизайн образовательных сайтов в период становления мобильного Интернета // Качество дистанционного

образования: концепции, проблемы, решения (DEQ - 2011). Межвузовский сборник научных трудов. М.: МГИУ. 2011. С. 123-136.

9. Каспаринский Ф.О. Использование программ ассоциативного картирования для управления распределенными информационными ресурсами // Научный сервис в сети Интернет: Труды XVII Всероссийской научной конференции. М.: ИПМ им.М.В.Келдыша. 2015. С.128-135.

# НАУЧНАЯ ИНФОРМАЦИОННАЯ СИСТЕМА СО ВСТРОЕННОЙ ТАКСОНОМИЕЙ НАУЧНЫХ ОТНОШЕНИЙ<sup>3</sup>

М.Р. Когаловский<sup>1</sup>, С.И. Паринов<sup>2</sup>

<sup>1</sup>Институт проблем рынка РАН, Москва

<sup>2</sup>Центральный экономико-математический институт РАН, Москва

**Аннотация.** Обсуждаются новые функциональные возможности научной информационной системы Соционет, основанные на использовании семантических связей между информационными объектами ее контента. Рассматриваются способы их создания, а также организация встроенной в систему таксономии семантических связей, определяющей их семантику.

## 1. Введение

Важной тенденцией развития технологий научных информационных систем (НИС) является поддержка семантики их контента и обеспечение на этой основе новых возможностей для пользователей. В данной работе акцентируется внимание на аспектах семантики, которые представляются семантическими связями между информационными объектами контента НИС, характеризующих прежде всего разнообразные научные отношения между ними. *Семантическими связями* мы называем бинарные ориентированные связи с *явным образом* определенной семантикой. Семантика связей определяется таксономией, встроенной в систему. Информационный объект, из которого исходит связь, называется далее ее *исходным объектом*. Объект, на который направлена связь, называется *целевым объектом* этой связи. Семантические связи представляются в системе как самостоятельные информационные объекты [1, 5]. Благодаря этому могут создаваться связи, в которых участниками также являются объекты-связи.

Разработке онтологий и таксономий семантических связей в контенте НИС посвящен ряд проектов в последние годы [1-4, 13-16]. В нашем случае семантические связи информационных объектов могут создаваться авторизованными пользователями децентрализованно в онлайн-режиме. В результате создания семантических связей формируется динамически обогащаемая *многослойная семантическая структура* контента, слои которой соответствуют различным классам созданных семантических связей [5, 6]. Анализ ее свойств позволяет решать ряд новых задач, наиболее важные из которых рассмотрены в разд. 4.

---

<sup>3</sup> Работа поддерживается РФФИ, проект 15-07-01294-а, и РГНФ, проект 14-02-12020-в

Обсуждаемая в данной работе технология семантического обогащения контента НИС с помощью декларации семантических связей реализована в системе *с открытым доступом* Соционет [12]. Это - крупная отечественная НИС, функционирующая более 15 лет и основанная на технологии открытых архивов [8]. Следует заметить, что обычно под *системой с открытым доступом*, функционирующей в среде Веб, понимается система, обеспечивающая свободное безвозмездное считывание и использование ее информационных ресурсов в некоммерческих целях. В нашем случае функция доступа трактуется более широко (аналогично трактовке его в технологиях баз данных) как возможность не только считывания, но и публикации в системе информационных объектов, в том числе, семантических связей, а также их актуализации, например, с помощью обновления метаданных.

В следующих разделах статьи обсуждаются организация встроенной в систему таксономии научных отношений между информационными объектами контента НИС, которую можно рассматривать и как *таксономию семантических связей* (разд. 2), способы создания семантических связей (разд. 3), а также функции системы, основанные на их использовании (разд. 4).

## **2. Организация таксономии семантических связей**

Декларация семантики связей в Соционет осуществляется пользователями при их создании с помощью встроенной в систему таксономии связей. Состав классов связей в таксономии продиктован стремлением обеспечить возможность создания в контенте системы семантических связей, позволяющих отображать разнообразные научные отношения между информационными объектами.

При разработке таксономии были использованы созданные в последние годы онтологии и таксономии связей. В частности, использованы фрагменты модульных комплексов онтологий *SPAR (The Semantic Publishing and Referencing Ontologies)* [14, 15] и *SWAN (Semantic Web Applications in Neuromedicine)* [13], рекомендации *SKOS (Simple Knowledge Organization System)* [16] консорциума W3C, проект *CRedit* [1, 4] открытого стандарта классификатора ролей участников коллективно выполняемого исследования. Используются также некоторые фрагменты модели научных данных *CERIF (Full Data Model)* [2], разработанной европейской организацией euroCRIS (<http://www.eurocris.org/>), а также онтологии [3], определяющей систему терминов, обозначающих сущности этой модели и отношения между ними.

Таксономия семантических связей, встроенная в Соционет, имеет двухуровневую структуру и организована в виде набора *контролируемых словарей*. Каждому классу верхнего уровня таксономии соответствует некоторый контролируемый словарь, и его подклассы (классы нижнего уровня) определяются значениями из этого словаря. Важно отметить, что семантические связи могут создаваться только для классов нижнего уровня, и



связь каждого класса может быть создана только между информационными объектами определенных типов.

Приведем для примера описание некоторых контролируемых словарей таксономии семантических связей, используемой в системе Соционет.

### **2.1. Словарь классов развития и дополнения результатов целевых публикаций**

Этот словарь включает классы связей, характеризующих варианты использования результатов целевых публикаций в исходных, в частности, для научного вывода новых результатов. Классы связей из этого словаря могут использоваться только для связей вида «публикация-публикация». Словарь включает следующие классы связей:

- Использует данные/метод/модель из
- Детализирует идею/метод/модель из
- Обобщает идею/метод/модель из
- Реализует идею/метод/модель из
- Анализирует результаты из
- Иллюстрирует результаты из
- Интерпретирует результаты из
- Исправляет ошибки в
- Уточняет результаты из
- Опровергает результаты из
- Развивает результаты из.

### **2.2. Словарь классов профессиональных оценок публикаций**

Словарь включает классы связей, характеризующих варианты оценки целевых публикаций со стороны пользователей системы, профили которых служат исходными объектами таких связей. Классы связей из этого словаря могут использоваться только для связей вида «персона-публикация». Этот словарь включает следующие классы связей:

- Инновационный результат
- Очень интересный результат
- Поворотный пункт для развития науки
- Наилучшая, наиболее релевантная работа по теме
- Оценивается позитивно
- Оценивается негативно
- Оценивается как ненаучная
- Содержит результаты, основанные на заблуждении
- Предполагается плагиат.

### **2.3. Словарь классов связей между компонентами и версиями публикаций**

Этот контролируемый словарь включает классы связей, характеризующих отношения между исходными участниками связей (абстракты, оглавления, предисловия, версии публикаций, презентации и др.) и

целевыми публикациями. Классы связей из этого словаря могут использоваться только для связей вида «публикация-публикация». Этот словарь включает следующие классы связей:

- Авторская версия (рукопись) для
- Версия с небольшими изменениями для
- Существенно переработанная версия для
- Пересмотренная или новая версия
- Идентичная копия для
- Презентация к
- Раздел/часть/глава из
- Абстракт (аннотация) для
- Оглавление для
- Предисловие (введение) из
- Список литературы из.

#### **2.4. Словарь классов мнений о существующих связях**

Классы связей этого словаря характеризуют мнения персон о семантике оценочных и других связей, выступающих в качестве целевых в данных связях. Персоны представляются в связи их профилями. Эти классы могут использоваться только для связей вида «персона-связь». К числу этих классов относятся:

- Готов улучшить свою публикацию
- Согласен со смыслом отношения
- Не согласен со смыслом отношения
- Неверно понята моя публикация
- Помогу улучшить эффект от использования моего результата
- Предлагаю подготовить совместную публикацию
- Предлагаю совместную работу по развитию моих результатов.

#### **2.5. Словарь классов полезных замечаний авторам публикаций**

Этот словарь включает классы связей публикаций создателя связи с рассматриваемой целевой публикацией. Семантика создаваемых связей предоставляет их авторам полезные сведения, основанные на содержании выбранных создателями связей их собственных публикаций (исходных публикаций связей). Классы связей из этого словаря могут использоваться только для связей вида «публикация-публикация». Словарь включает следующие классы:

- Ваши результаты анализируются в моей публикации
- Ваши идея/метод/модель/результаты обобщаются в моей публикации
- Ваши идея/метод/модель/результаты детализируются в моей публикации
- Ваши идея/метод/модель/результаты иллюстрируются в моей публикации
- Ваши идея/метод/модель/результаты реализованы в моей работе
- Ваши результаты интерпретируются в моей публикации
- Данные/метод/модель в моей публикации лучше

В моей публикации обсуждается близкая проблема  
В моей публикации опровергаются ваши результаты  
В моей публикации получены те же результаты  
В моей публикации отмечены и исправлены Ваши ошибки.

## **2.6. Словарь видов вкладов соавторов в создание коллективных публикаций**

Словарь включает классы связей, характеризующих вклад персон-соавторов в подготовку коллективных публикаций - целевых объектов связей. При этом профиль персоны-соавтора выступает в качестве исходного участника связи. Классы связей из этого словаря могут использоваться только для связей вида «персона-публикация». К ним относятся классы из таксономии *CreDit* [1]:

- Подготовка начального варианта текста
- Визуализация данных, подготовка презентации
- Критический анализ/комментирование или доработка текста
- Проведение экспериментов
- Разработка методологии
- Исследование концепции
- Сбор данных, аргументов
- Компьютерные работы
- Обеспечение ресурсами
- Формальный анализ
- Курирование данных
- Администрирование проекта
- Научное руководство
- Проведение исследования
- Концептуализация проблемы
- Обеспечение финансирования.

Набор контролируемых словарей семантических связей в системе Соционет может пополняться новыми словарями, словари могут дополняться новыми классами, и имена классов в словарях могут изменяться.

## **3. Способы создания семантических связей**

Возможность создания семантических связей в системе Соционет доступна только для ее *авторизованных пользователей*. Идентификация пользователя необходима системе для указания его авторства в описаниях создаваемых им связей, а также для определения доступных ему функций в зависимости от того, является ли он автором или соавтором рассматриваемой публикации. Идентификация создателя связи необходима и потому, что факт существования связи и ее семантика становятся в системе публично доступными, и отсутствие анонимности автора повышает его ответственность

перед научным сообществом за мнение или оценку, представляемые создаваемой им связью.

В Соционет реализованы несколько способов создания семантических связей. Основной способ - использование *специального интерфейса* на страницах описаний публикаций, которые станут целевыми участниками создаваемых связей, и возможен как для автора данной публикации, так и для ее читателя. Поведение системы и доступные пользователю возможности зависят при этом от того, является ли он автором или читателем этой публикации. Для *пользователя-автора* исходным объектом создаваемой связи может быть его персональный профиль, а целевым объектом – данная его публикация. Он может указать свой вклад в подготовку этой публикации с помощью словаря 2.6, если она – коллективная, может аннотировать абстракт публикации с целью актуализации или уточнения его содержания. Кроме того, пользователь-автор может установить семантическую связь между парой своих публикаций, используя словарь 2.1, указывая отношение между изложенными в них результатами, либо указать с помощью словаря 2.3 характер отношений между различными версиями или компонентами одной и той же своей публикации.

*Пользователь-читатель* может установить связь от своего персонального профиля (исходный объект) к этой публикации другого автора (целевой объект). Он может аннотировать фрагменты ее абстракта, дать профессиональную оценку публикации с помощью словаря 2.2, может создать связь от выбранной своей публикации (исходный объект) к этой чужой публикации, например, определяемую словарем 2.5 или 2.1.

Еще один способ создания семантических связей в системе Соционет - семантическое обогащение связей цитирования - применяется к связям между данной публикацией как исходной и публикациями, указанными в пристатейном списке литературы [9]. Фактически, речь идет о семантическом обогащении «немых» связей цитирования, представленных в описании публикации. «*Немыми*» мы называем традиционные связи цитирования, поскольку они не несут какой-либо информации о мотивах, побудивших автора цитировать ту или иную публикацию. Для таких связей в результате определяется явно указанная семантика с помощью словаря 2.1. Такие связи может создавать только *пользователь-автор цитирующей публикации*.

Как уже упоминалось, целевым участником связи может быть ранее созданная связь. Вызвав описание целевой связи, пользователь может декларировать свое мнение о ней, создавая новую связь своего профиля с этой связью с использованием словаря 2.4.

#### **4. Функции системы, основанные на использовании семантических связей**

Обработка информации о семантических связях между информационными объектами, определенных в Соционет, обеспечивает ряд новых возможностей системы.

Одна из ключевых - обеспечение *новых форм научных коммуникаций* в сообществе пользователей Соционет. Если пользователем создана новая семантическая связь или модифицировано описание уже существующей связи, в системе активизируется *сервис уведомления*. Этот сервис генерирует необходимую информацию об этом событии, которая показывается авторизованному пользователю на страницах системы Соционет, а также в некоторых случаях может быть направлена в виде сообщения по электронной почте авторам связываемых информационных объектов. Адреса электронной почты автора связи и адресатов сообщения известны системе из их профилей.

Публичная доступность созданных семантических связей и передаваемых с их помощью сообщений обуславливает ответственное отношение к ним породивших их пользователей. Эти сообщения могут носить оценочный характер – своего рода рецензий целевой публикации связи. Они могут информировать авторов публикаций о близких к ним работам, в которых используются и каким образом результаты автора, либо получены аналогичные результаты, могут нести полезную информацию авторам публикаций по развитию их работ. Получение сообщения *стимулирует адресата* реагировать на него, создавая соответствующую новую связь, и тем самым поддерживать научную дискуссию. Более подробно коммуникационные возможности Соционет обсуждаются в [7].

Сценарии функционирования сервиса уведомления управляются рядом параметров, которые могут специфицироваться авторами информационных объектов контента системы и пользователями, действия которых инициируют передачу им сообщений. Автор информационных объектов может, например, установить в своем профиле блокировку передачи ему сообщений, а автор создаваемой связи указать, что сообщение нужно или не нужно передавать. Если участником создаваемой или модифицируемой связи является коллективная публикация, то он может указать, всем или какому-либо конкретному соавтору следует передать сообщение.

Другая функция системы, основанная на использовании семантических связей – их *контекстная визуализация* и обеспечение пользователям доступа к информационным объектам контента путем *семантической навигации* по структуре связей. При обращении пользователя к какой-либо из публикаций он имеет возможность визуально анализировать ближайшую окрестность соответствующего ей узла семантической структуры, т.е. все непосредственно связанные с ним узлы и классы всех входящих и исходящих его связей. Пользователь может также осуществлять навигацию по сети, пошагово перемещаясь от одной публикации к другой по связям интересующих его классов. При этом можно отфильтровать слой сети, соответствующий интересующему классу связей. Навигацией по полученной подсети можно проследить, например, эволюцию научных результатов в некоторой области науки, выявить относящиеся к ней основополагающие работы. Используя

подсеть оценочных связей, можно найти работы, в которых данная публикация подвергается критике или, наоборот, оценивается позитивно.

*Аннотирование фрагментов абстрактов* публикаций – еще одна функция, основанная на использовании семантических связей. Она представляется как автору для актуализации его представления о проблеме, обсуждаемой в данной публикации, так и другим пользователям системы.

Следует далее упомянуть *поддержку «живых» документов*. Авторы некоторых электронных публикаций продолжают работать над ними и заинтересованы в поддержке их содержания в актуальном состоянии. Электронные документы, содержание которых корректируется на протяжении времени, называются *«живыми»* документами [10]. Для использования «живых» документов средствами обсуждаемых здесь технологий недостаточно указывать в их заголовках даты новых версий или их номера, поскольку такие документы могут быть участниками семантических связей, которые, например, характеризуют их оценку. Эта оценка может стать неадекватной после внесения автором изменений в документ. Решение проблемы заключается в том, что при внесении изменений в информационный объект–участник связи автор связи получает уведомление об этом событии и может при необходимости внести соответствующее изменение в описание этой связи или удалить ее. Актуализация «живого» документа может осуществляться также с помощью аннотирования фрагментов его абстракта.

Отметим, наконец, еще одну важную функцию системы Соционет, основанную на использовании семантических связей, - *наукометрические исследования* контента системы. Динамически пополняемая и обновляемая его семантическая структура может служить источником данных для наукометрических исследований, более информативным по сравнению с традиционно используемыми [11]. Обычно наукометрические исследования основаны на использовании связей цитирования в пристатейных списках публикаций из периодики. Однако эти связи не несут информации о мотивах цитирования в отличие от семантически обогащенных связей цитирования в системе Соционет. Кроме того, для наукометрии в системе могут использоваться и поддерживаемые в системе связи, отличные от связей цитирования, имеющие различную семантику.

Наукометрические сервисы Соционет генерируют статистические данные *дифференцировано по классам связей*, в том числе, и для связей цитирования. Такие данные позволяют более глубоко анализировать и оценивать научное качество представленных в системе публикаций, степень их влияния на развитие научных исследований, изучать наукометрические «портреты» как отдельных ученых, так и исследовательских организаций в целом, получать более осмысленные оценки их научной продуктивности. Использование семантических связей, отражающих вклады авторов в создание коллективных публикаций, позволяет также получать персональные оценки этого аспекта

научной деятельности отдельных исследователей и организаций, с которыми они аффилированы.

## **Заключение**

Работа по развитию обсуждаемой в статье технологии продолжается. Ведется также работа по интернационализации системы Соционет.

## **ЛИТЕРАТУРА**

1. Allen, L., Brand, A., Scott, J., Altman, M., and Hlava, M.: Credit where credit is due. *Nature/ International weekly journal of science. Nature*. Vol. 508, Issue 7496 (April 2014). URL: [http://www.nature.com/polopoly\\_fs/1.15033!/menu/main/topColumns/topLeftColumn/pdf/508312a.pdf](http://www.nature.com/polopoly_fs/1.15033!/menu/main/topColumns/topLeftColumn/pdf/508312a.pdf)
2. CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS, 2012. URL: [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3\\_FDM.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf)
3. CERIF 1.3 Semantics: Research Vocabulary. CERIF Task Group, euroCRIS, 2012. URL: [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3\\_Semantics.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf)
4. CRedit. An open standard for expressing roles intrinsic to research. Taxonomy. URL: <http://credit.casrai.org/proposed-taxonomy/>
5. Когаловский М.Р., Паринов С.И. Семантическое структурирование контента научных электронных библиотек на основе онтологий. В кн.: "Современные технологии интеграции информационных ресурсов: сборник научных трудов". Санкт-Петербург: Президентская библиотека им. Б.Н. Ельцина, 2011. - С. 26-45.
6. Kogalovsky, M.R., Parinov, S.I.: Social Network Technologies for Semantic Linking of Information Objects in Scientific Digital Library. *Programming and Computer Software*. Vol. 40, No. 6, 313–320 (2014).
7. Когаловский М.Р., Паринов С.И. Научные коммуникации в среде семантически обогащаемых электронных библиотек //Программная инженерия. - 2015. - № 4. - С. 31-38.
8. Open Archives Initiative. URL: <http://www.openarchives.org/>
9. Parinov, S.: Semantic enrichment of research outputs metadata: new CRIS facilities for authors. *Proc. of MTSR 2014, 8<sup>th</sup> Metadata and Semantics Research Conference, 27-29 November 2014, Karlsruhe, Germany (2014)*.
10. Паринов С.И., Когаловский М.Р. «Живые» документы в электронных библиотеках //Прикладная информатика. - 2009. - № 6 (24). - С. 123-131.
11. Parinov, S., Kogalovsky, M.: Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. *Scientometrics*. Vol. 98, Issue 2, 927-943 (2014).
12. Паринов С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов //Российский научный электронный журнал

- «Электронные библиотеки». - 2003. - Том 6. - Вып. 1. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP>
13. Semantic Web Applications in Neuromedicine (SWAN) Ontology. W3C Interest Group Note, 20 October 2009. <http://www.w3.org/TR/hcls-swan/>
  14. Shotton, D.: Open Citations and Related Work. Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. URL: <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/>
  15. Shotton, D., Peroni, S.: Semantic annotation of publication entities using the SPAR (Semantic Publishing and Referencing) Ontologies /Beyond the PDF Workshop, La Jolla, 19 January 2011. URL: <http://speroni.web.cs.unibo.it/publications/shotton-2010-semantic-annotation-publication.pdf>
  16. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 August 2009. URL: <http://www.w3.org/TR/skos-reference/>



# АВТОМАТИЗАЦИЯ СОСТАВЛЕНИЯ ОТЧЕТОВ О НАУЧНОЙ ДЕЯТЕЛЬНОСТИ

А.С. Козицын, С.А. Афонин, Д.Д. Голомазов, М.А. Занчурин, И.С. Астапов

*НИИ Механики МГУ им. М.В. Ломоносова*

Управление любой организацией должно опираться на использование актуальных данных о её состоянии организации и о качестве выпускаемой продукции. Если для большинства промышленных предприятий оценка количества и качества продукции не представляет сложностей, то оценка деятельности научно-образовательных организаций и выпускаемой ими продукции является сложной задачей, как в организационном, так и в техническом плане. Существует большое число разрозненных критериев оценки, таких как индексы цитируемости, финансовые показатели выполняемых проектов, доклады на конференциях. Однако, сбор и обработка этой информации традиционными средствами слишком трудоемка. В этой связи, во многих научных организациях поднимается вопрос об использовании для решения этих задач специализированных автоматизированных систем сбора, обработки и представления данных о научной деятельности сотрудников.

В Московском государственном университете имени М.В. Ломоносова несколько лет назад внедрена и успешно используется система «Наука МГУ», созданная на основе Интеллектуальной Системы Тематического Исследования Научно-технической информации - "ИСТИНА" (далее - Система)[1][2]. Система осуществляет сбор и анализ информации о публикуемых научными сотрудниками статьях в журналах и сборниках, о монографиях, докладах на конференциях, грантах, проектах, полученных наградах, участии в различных оргкомитетах и редколлегиях, о ведении учебно-педагогической работы и получении прав на интеллектуальную собственность. Автоматическая оценка цитируемости публикаций позволяет определить рейтинг публикация, отдельных авторов и подразделений в целом. Такая оценка используется при организации и проведении различного рода конкурсов.

Собранная в Системе информация используется для создания личных страниц сотрудников, построения аналитических отчетов, а также для автоматизированного построения отчетных форм о научной деятельности.

Отображение основных достижений и результатов научной деятельности сотрудника на его личной странице не только популяризует его научные результаты, но и способствует увеличению цитируемости его работ. В настоящее время большую часть информации люди получают из Интернет, в том числе при подборе материалов по тематике своей научной работы. Предоставление возможности Интернет-поиска по своим работам и простого

интерфейса для получения библиографических данных по ним позволяет повысить личный индекс цитирования. Данный аспект имеет немаловажное значение, как для каждого сотрудника, так и для организации в целом, поскольку мировые рейтинги ВУЗов (например, ТНЕ, QS и Шанхайский) составляются с учетом цитируемости публикаций их сотрудников и представления информации об этих ВУЗах в Интернет. Помимо этого представление качественной персональной страницы с научными результатами повышает общий рейтинг научного сотрудника при первичном знакомстве с ним потенциальных партнеров, экспертов по оценке тех или иных заявок, редакторов журналов и других людей.

Аналитические отчеты позволяют на основе таблиц и диаграмм просматривать сводную информацию за выбранный период как по всем, так и по отдельным, заранее выбранным подразделениям, с разбивкой по тематическим каталогам ГРНТИ и Scopus. Анализироваться могут следующие показатели: число статей, число авторов, число ссылок по Web of Science и другие (Рис. 1).

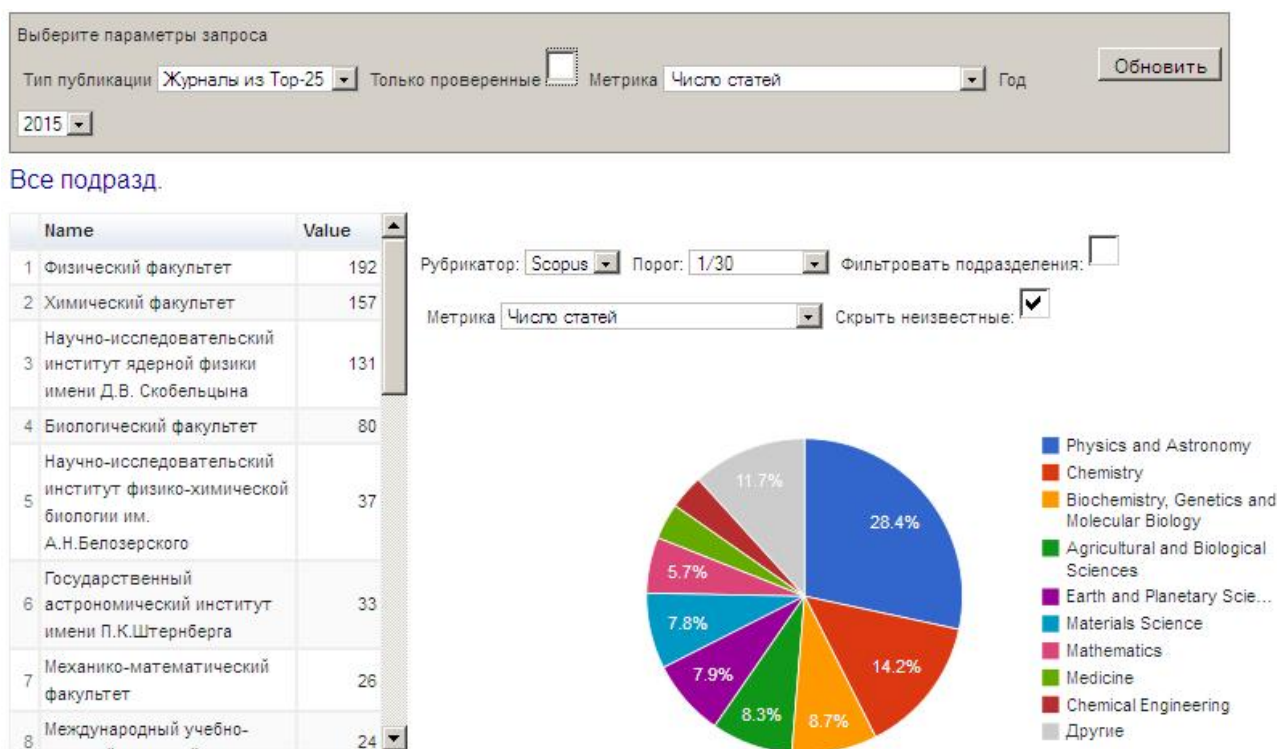


Рис 1. Пример аналитического отчета.

Построение стандартных отчетных форм осуществляется на основе разработанного авторами генератора отчетов [3]. На настоящее время разработанные отчетные формы позволяют в автоматическом режиме создавать отчеты по подразделениям о состоянии выполнения научно-исследовательских работ и их финансировании, о премиях, публикациях, конференциях, о

результатах патентной деятельности, об учебной работе со студентами и послевузовской подготовки кадров, о защите диссертаций, взаимодействии с зарубежными организациями, о стажировках и участии в деятельности зарубежных сообществ. На основе этих данных могут составляться таблицы итоговых индикаторов эффективности работы подразделений. Помимо подготовки отчетов подразделений, Система позволяет в автоматическом режиме составлять персональные отчеты сотрудника за тот или иной период времени (за год, за 5 лет и т.д.), его информационный лист, списки статей из Web of Science и Scopus, список статей из перечня журналов, представленных в списке ВАК, а также другие материалы, необходимые сотрудникам для подготовки пакета документов при прохождении различных аттестаций, конкурсов и других формальных мероприятий.

Подобная автоматизация процесса сбора информации для подготовки отчетности имеет несколько преимуществ. Во-первых, использование одних и тех же данных экономит время сотрудников и делает отчеты непротиворечивыми. Информацию о каждой работе требуется внести в Систему только один раз, после этого информация отображается сразу во всех отчетных формах. Во-вторых, информация верифицируема. При построении итоговых отчетных форм есть возможность посмотреть детальную информацию по любым итоговым данным, нажав на указанное в отчете число (Рис. 2).

Таб. 2.2 Научные публикации научно-педагогических и научных работников  
Количество монографий указывается без учета учебных пособий.

Summary Table Data:

Всего моногр.	Моногр. (РФ)	Моногр. выполненных штат. преподавателями	Объем монографий	Объем Моногр. выполненных штат. преподавателями	В сборн МГУ	В сборн (не МГУ)	Главы	Статей в журналах РФ
1	1	0	16,38	0	0	3	0	13

Journal Articles Table:

ID	Авторы	Название	Журнал	Номер	Год
7675272	Шундеев А.С.	Система распределенных вычислений на базе платформы Erlang/OTP	Программная инженерия	5	2014
7675295	Шундеев А.С.	Виртуальный компьютерный класс	Программная инженерия	7	2014
5949505	Васенин В.А., Роганов В.А., Зензинов А.А.	Среда моделирования для исследования средств обеспечения информационной безопасности в Grid и Cloud - системах	Программная инженерия	3	2014
5679587	Афонин С.А., Гаспаряц А.Э.	Разрешение неоднозначности авторства публикаций при автоматической обработке библиографических данных	Программная инженерия	1	2014
5948586	Васенин В.А., Афонин С.А., Панюшкин Д.С.	Модели распространения информации в социальных сетях	Программная инженерия	2	2014
6222686	Васенин В.А., Афонин С.А., Панюшкин Д.С.	Модели распространения информации в социальных сетях	Программная инженерия	4	2014

Рис. 2. Детализация итоговых отчетных данных.

Еще одним достоинством разработанного подхода является возможность создания пользователями собственных отчетов на основе требований и регламентов отдельных подразделений или организаций [4]. Создание собственных отчетных форм не требует внесения изменений в программный код системы или знания особенностей ее реализации. Она может производиться как на основе уже существующих отчетных форм так и по новым, не имеющим аналогов, направлениям.

Вместе с тем, создание сложных отчетных форм и предоставление пользователям широких возможностей по разработке новых отчетов создает определенные риски в области обеспечения отказоустойчивости Системы. Выполнение большого числа сложных отчетов может вызывать перегрузку сервера Базы Данных (БД) и замедление работы Системы в целом. С целью предотвращения подобных ситуаций в архитектуре разработанной Системы предусмотрено распараллеливание нагрузки по доступу к БД. Основной сервер БД обслуживает запросы только основного ядра Системы, а обработку запросов отчетных форм пользователей осуществляют дополнительные сервера БД, которые синхронизируются с основным сервером [5]. На настоящий момент в Системе используется два дополнительных сервера. Созданная архитектура позволяет добавлять такие сервера для обеспечения масштабируемости Системы практически не ограничено.

Использование разработанной Системы при подготовке отчетов организации за 2014 год и проведении конкурсов показало эффективность предложенных решений.

## ЛИТЕРАТУРА

1. Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА) / С.А.Афонин и др. Под ред. академика В.А.Садовниченко. М.: Издательство Московского университета. 2014. – 262 с.
2. Система ИСТИНА. — URL: [istina.msu.ru](http://istina.msu.ru)
3. Sergey A., Alexander K., Ivan A. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. - 2014. - P. 529-534.
4. Астапов И., Афонин С., Козицын А. Формирование и использование настраиваемых отчетов в системе ИСТИНА // Ломоносовские чтения. Научная конференция. Секция механики. 14-23 апреля 2014 года. Тезисы докладов. - Издательство Московского университета Москва, 2014. - С. 21-22.
5. Козицын А., Афонин С., Голомазов Д. Организация хранения и доступа к данным в информационных системах на примере системы ИСТИНА // Ломоносовские чтения. Научная конференция. Секция механики. 14-23 апреля 2014 года. Тезисы докладов. - Издательство Московского университета Москва, 2014. - С. 85-86.

# АЛГОРИТМ СЛИЯНИЯ ДУБЛЕТНЫХ БИБЛИОГРАФИЧЕСКИХ ЗАПИСЕЙ

К.А. Косолапов<sup>1</sup>, В.А. Серебряков<sup>2</sup>, К.Б. Теймуразов<sup>2</sup>, О.Н. Шорин<sup>3</sup>

*1 Московский государственный университет им. М.В. Ломоносова*

*2 ФГБУН Вычислительный центр им. А.А. Дородницына Российской академии наук*

*3 ФГБУ «Российская национальная библиотека»*

## Абстракт

В ходе реализации системы публикации библиографических записей с использованием технологий семантической паутины возникла задача создания эффективного алгоритма по поиску нечетких дубликатов среди большого количества записей. В работе рассматривается использование функции из семейства locality-sensitive hashing с дополнительными оптимизациями для выявления дублетных библиографических записей. Выявленные таким образом записи сравниваются друг с другом с помощью меры Жаккара для принятия окончательного решения о дублетности и осуществления слияния.

## Введение

В Российской государственной библиотеке и Российской национальной библиотеке реализуется проект по публикации библиографических записей с использованием технологий семантической паутины. Основной целью проекта является создание программной системы, которая бы в автоматическом режиме собирала библиографические записи из различных библиотек, осуществляла бы связывание данных и публиковала бы их в Linked Open Data. Наличие открытого доступа к крупнейшему массиву данных, ориентированному для использования не только человеком, но и автоматизированными средствами, позволит создать новые высокоинтеллектуальные онлайн-сервисы, которые окажут значительное влияние на развитие культуры и книжной отрасли.

Описываемая система создается как часть Национальной электронной библиотеки (НЭБ), которая обеспечит свободный, равный и всеобщий доступ граждан нашей страны к документной информации историко-культурного, научного и образовательного назначения через сеть Интернет, предоставляемый на основе единой общенациональной системы создания и эффективного использования цифровых библиотечно-информационных ресурсов и сервисов. Основным отличием от проектов подобного рода, например, сводного каталога электронных ресурсов (СКЭР) является наличие единого интерфейса для работы с документами из различных библиотек. Предыдущие проекты предоставляли только ссылки на электронные ресурсы, для получения доступа к которым необходимо было регистрироваться в каждой

конкретной библиотеке, устанавливать и настраивать специализированное программное обеспечение, применяемое в данной библиотеке.

Для создания модульной системы публикации данных, способной без значительных усилий подключать новых участников, работы были разбиты на несколько этапов [2]:

- 1) Разработка онтологии предметной области на базе существующих решений.
- 2) Осуществление интеграции с автоматизированными библиотечными информационными системами.
- 3) Осуществление конвертации библиографических записей из форматов MARC21 и Rusmarc в унифицированный формат - MODS.
- 4) Решение вопроса о хранении сконвертированных данных.
- 5) Выявление дублетных записей, полученных из различных библиотек, и осуществление взаимного обогащения данных из этих записей.
- 6) Выбор данных для связывания и публикации данных в Linked Open Data.
- 7) Реализация модуля визуализации полученного результата.

Данная система имеет распределенную структуру: библиотеки выступают поставщиками библиографических записей, которые аккумулируются, хранятся и обрабатываются на центральном сервере. В мире существует ряд проектов, направленных на интеграцию библиотечных данных и использующих схему аккумуляции данных на центральном сервере из распределенных источников. Среди них можно выделить такие проекты как сводный каталог библиотек России (СКБР), объединенный мировой каталог WorldCat, европейская цифровая библиотека Europeana. Тщательный анализ используемых в этих проектах решений является залогом успешного функционирования сервиса по семантической интеграции библиографических записей.

### **Постановка задачи**

Выполняя свои уставные функции, библиотеки создают библиографические записи на экземпляры, хранящиеся в их фондах. Общее количество записей, хранящихся только в двух крупнейших библиотеках страны – РГБ и РНБ, составляет несколько десятков миллионов. Поскольку обе эти библиотеки являются получателями обязательных экземпляров, то большая часть их фондов совпадает. Как следствие, в библиотечных системах хранятся записи на одни и те же произведения – дублетные библиографические записи. Важно отметить, что в РНБ для описания книг используется формат, разрабатываемый и поддерживаемый Национальной службой развития системы форматов – Rusmarc, а в РГБ используется американский формат описания произведений – MARC21. Нельзя не упомянуть, что в этих библиотеках применяются различные стандарты на полноту заполнения описания – какие поля являются обязательными, а какие – дополнительными, а также на строение отдельных полей.

Стоит подчеркнуть, что сводный каталог библиотек России (СКБР) является частью каталога НЭБ. Также в каталог НЭБ попадают записи из региональных библиотек, которые зачастую являются участниками СКБР. Таким образом, можно утверждать, что количество дублетных библиографических записей, попавших в каталог НЭБ, велико.

Для выявления дублетных библиографических записей, полученных из разных источников и заполненных с использованием различных стандартов, требуется создание эффективного алгоритма. При получении нескольких десятков миллионов записей из нескольких библиотек задача сравнения всех записей со всеми для выявления дублетных записей на одни и те же книги становится неприемлемо затратной –  $O(n^2)$ . Возникает необходимость создания алгоритма, позволяющего существенным образом сузить множество библиографических записей – потенциальных кандидатов на дублетность. Большинство подобных алгоритмов основано на разбиении всего множества данных на кластеры, внутри которых содержатся подобные друг другу записи. Необходимо проанализировать существующие алгоритмы выявления записей на один и тот же объект с учетом присутствия ошибок и аббревиатур, различий стандартов заполнения полей.

### **Обзор существующих решений**

Априорно известно, что при получении библиографических записей из разных библиотек, часто будет возникать ситуация, когда на одну и ту же сущность будет существовать несколько записей. Эти записи будут отличаться как по формату, так и по полноте заполнения, поскольку в различных учреждениях процессы каталогизации отличаются друг от друга. Например, состав дополнительных элементов, точек доступа может быть разным, системы классификации и предметизации, использование аббревиатур также могут различаться от одной организации к другой. К тому же, запись может содержать ошибки, опечатки, вызванные банальной человеческой оплошностью и невнимательностью.

Вопросы интеграции библиографических записей из различных источников с последующим их объединением и обогащением давно находятся в фокусе внимания ученых. Одними из основоположников этого направления являются Ivan P. Feleggi и Alan B. Sunter [4], которые разработали математическую модель, позволяющую разделить множество записей на несколько кластеров. В кластер попадают записи, которые в терминах той или иной метрики располагаются недалеко друг от друга. Для выявления дублетных записей достаточно сравнить записи, входящие в состав одного кластера, что значительно снижает количество сравнений.

Jeremy A. Hylton [5] развил идеи Feleggi и Sunter, распространив их на проблему не только выявления дублетных записей, но и создания на их основе обогащенной записи, содержащей объединенную информацию из нескольких записей, с последующим удалением дублетных записей, содержащих менее

полную информацию. В его работе показано, что процесс разбиения записей на кластеры должен предваряться процедурой нормализации – набором правил, применение которых приводит библиографические записи к некоему единообразному виду. К таким правилам нормализации можно отнести удаление избыточных пробельных символов, приведение строк к одному регистру, замена общепризнанных аббревиатур и правил написания на единообразный. Например, даты, записанные в разных форматах: Sept. 1987, 09.1987 – заменялись на какой-то один формат.

Поскольку зачастую дополнительная информация, указанная в записи в виде ссылок, может быть недоступна, например, получить полный текст произведения по ссылке из описания невозможно из-за ограничений, накладываемых авторским правом, то процесс выявления дублетных записей может ориентироваться только на информацию, содержащейся в самой записи. Таким образом, целый класс алгоритмов, использующих дополнительную информацию, не может быть нами использован. Мы можем использовать только основные поля записи – «Автор» и «Название», т.е. по сути оперировать строками.

Одними из наиболее распространенных метрик для расчета расстояния между строками являются меры Хемминга, Евклида, Левенштейна, Джаро-Винклера, Рэтклиффа-Обершелпа. В работе Д.Н.Рубцова и В.К.Баракнина [1] показано, что для простейшего случая, когда сравнение записей осуществляется только по полям «Автор» и «Название», достаточно использование одного из методов динамического программирования, предложенного Хиршбергом. Данный метод обладает высокой эффективностью и относительно простой реализацией.

Также в вышеупомянутой работе описан ряд исключений, когда несколько записей при формальной практически полной идентичности содержат информацию о различных объектах. В частности, к таким исключениям можно отнести описания на отдельные тома многотомных изданий.

Выбор данного решения был обусловлен тем, что количество библиографических записей в используемой системе не столь велико. В нашем же случае количество библиографических записей составляет несколько десятков миллионов, что делает невозможным оперирование строками, так как сложность описанного алгоритма составляет  $O(mn)$ , где  $m$  – максимальная длина строки, а  $n$  – число уже имеющихся записей.

### **Алгоритм слияния дублетных библиографических записей**

В отличие от строк, поиск и сравнение среди большого количества натуральных чисел можно организовать более эффективно. Достаточно разместить их в отсортированном массиве. Тогда, используя алгоритм двоичного поиска, можно найти необходимое число за  $O(\log n)$  операций сравнения.



Для перевода строковых значений в числовые используются функции хэширования. Однако, большинство алгоритмов хэширования не подходят для нашей задачи, так как изначально они создавались для равномерного отображения пространства строк на пространство чисел. Это означает, что при незначительном изменении аргумента функции хэширования – строки, значение функции – число может измениться разительно. В нашем же случае необходимо добиться обратного результата – незначительное изменение аргумента должно приводить к незначительному изменению результата функции хэширования.

Подобные алгоритмы хэширования составляют семейство locality-sensitive hashing (LSH) функций. Одно из определений семейства таких функций дано в работе P. Indyk и R. Motwani [6]. Семейство  $F$  функций  $h: M \rightarrow S$  называется  $(R, cR, P_1, P_2)$ -чувствительным, если для порога  $R > 0$ , коэффициента  $c > 1$  и любых двух точек  $p, q \in M$  выполнены два условия:

1. Если расстояние  $d(p, q) \leq R$ , то  $h(p) = h(q)$  с вероятностью не меньше  $P_1$ .
2. Если расстояние  $d(p, q) \geq cR$ , то  $h(p) \neq h(q)$  с вероятностью не более  $P_2$ .

Альтернативное определение семейства таких функций дано в работе M. Charikar [3]: для  $\text{sim}$  – функции подобия объектов из множества  $P$ :  $\text{sim}: P \times P \rightarrow [0, 1]$ , то схема LSH – это семейство хэш-функций  $H$ , имеющих распределение  $D$ , таких что при выборе функции  $h \in H$  согласно распределению  $D$ :

$$\Pr_{h \in H}[h(q) = h(p)] = \text{sim}(q, p), \forall q, p \in P$$

Одним из наиболее часто используемых представителей из семейства LSH-функций является функция SimHash. Шаги алгоритма, реализующего данную функцию, следующие:

1. Исходная строка разбивается на слова, между которыми стоят разделительные знаки (пробелы, знаки отступа, перенос строки и т.д.), в результате чего получается массив строк.
2. Создаётся массив целых чисел, заполненный нулями, с размером, равным длине хеша в битах.
3. Для каждого слова вычисляется хэш-функция в виде:  

$$\text{hash}(s) = s[0] * 31^{n-1} + s[1] * 31^{n-2} + \dots + s[n-1]$$
4. Для каждого бита, полученного хеша, соответствующий ему элемент массива изменяется следующим образом: элемент массива увеличивается на единицу в случае, если исходный бит равен 1 и уменьшается на единицу в противном случае.
5. На основе полученного массива генерируется результат хэширования – последовательность бит (в данном случае 32 бита, представленные типом int). Используется следующее соответствие элементов массива и хеша: если элемент массива больше нуля, то соответствующий бит равняется единице, иначе нулю.

Как можно заметить, данная схема хеширования не зависит от порядка слов в исходном наборе данных, например, «синтаксис и семантика» и «семантика и синтаксис» будут иметь одинаковые хэши.

Поскольку количество записей велико, то сгенерировав хэши для всех записей, появляется проблема эффективного поиска среди этого набора, поскольку в худшем случае для поиска близких хэшей необходимо произвести  $2^{32}$  операций сравнения. Для снижения количества сравнений был использован метод разбиения хэша на равные части, описанный в работе Bingfeng Pi, Shunkai Fu, Weilei Wang и Song Han [7]. В данной работе предлагается делить полученный хэш  $hash$  на равные части  $hash_1, hash_2, \dots, hash_n$ . Тогда поиск близких хэшей можно осуществить с некоторой оптимизацией: среди уже имеющихся хэшей находятся такие, в которых совпадает хотя бы одна из частей. При разбиения хэша на 4 равные части подобная оптимизация позволяет найти близкие хэша в худшем случае за  $4 \cdot 2^{24}$  операции сравнения.

Сделав предположение о том, что различие двух близких хэшей будет локализовано в половине частей, можно добиться дополнительной оптимизации. Если выбирать хэши, в которых совпадает только  $n/2$  из  $n$  частей, то при  $n$  равном 4 в худшем случае необходимо будет сделать  $6 \cdot 2^{16}$  операций сравнения. Таким образом, для хэша, имеющего вид  $hash_1$ - $hash_2$ - $hash_3$ - $hash_4$ , найдутся все записи, в которой первая и вторая части равны  $hash_1$ - $hash_2$ , первая и третья части равны  $hash_1$ - $hash_3$  и т.д.

Таким образом, используя вышеописанный алгоритм для поиска близких хэшей для полей «Автор» и «Название», мы отберем множество записей – претендентов на дублетность. Для каждого претендента мы осуществляем сравнение строк с использованием меры Жаккара:  $J(A, B) = |A \cap B| / |A \cup B|$  - соотношение количества совпадающих элементов к объединению. Возникает вопрос о том, что взять за единичный элемент, из которого состоят строки. На первый взгляд кажется, что наиболее логичным решением является разделение на отдельные слова.

Однако в случае, если слова в исходных строках будут близки, но не идентичны, например, «воздушный поток» и «возд. поток», то близость слов «воздушный» и «возд.» будет утеряна. Чтобы это предотвратить, исходной строке ставится в соответствие неупорядоченный набор признаков. Существует несколько подходов для создания таких признаков. Самым распространенным является поиск  $n$ -грамм.

$N$ -грамма – последовательность  $n$  символов или слов. Например, для словосочетания «синтаксис и семантика» биграммами ( $n$ -грамма для  $n$  равного 2) с точки зрения слов являются «синтаксис и», «и семантика». С точки зрения символов все биграммы этой строки это: «си», «ин», «нт», «та», «ак», «кс», «си», «ис», «с », « и», «и », « с», «се», «ем», «ма», «ан», «нт», «ти», «ик», «ка». За счет того, что в исходной задаче требуется идентификация небольших строк, располагающихся в полях библиографических записей, имеет смысл использовать подход, трактующий биграммы с точки зрения символов,

поскольку на небольших по объему данных он генерирует наборы большого размера.

Для записей, претендующих на совпадение, посчитаем меры Жаккара для множеств биграмм, полученных из полей «Автор» и «Название», и возьмем среднее арифметическое. Если полученное значение будет выше заданной планки, а указанные записи не подпадают под исключения, описанные в работе Д.Н.Рубцова и В.К.Барахнина [1], то установим соответствие между этими записями.

Текущая реализация алгоритма выявления дублетных библиографических записей обрабатывает 10 тыс. новых записей на хранящихся в РГБ 15 млн. записях за 34 секунды. В РНБ хранится около 10 млн. библиографических записей. Таким образом, выявление всех дублетных записей РНБ и РГБ займет 34 тыс. секунд, т.е. около девяти с половиной часов. В процессе своей работы текущая реализация алгоритма создает дубликаты всех оригинальных записей, а также хранит таблицы хэшей для авторов и названий. Для хранения хэшей 25 млн. записей требуется  $32 \cdot 2 \cdot 25$  млн. бит, что приблизительно равняется 190 Мб. Место, необходимое для хранения оригинальных записей, напрямую зависит от количества найденных дублетов в массивах РГБ и РНБ. По оценкам экспертов около 70% записей в РНБ и РГБ являются дублетами. Таким образом, количество оригинальных записей составляет 18 млн. Т.к. в среднем библиографическая запись занимает 1 Кб, то для хранения оригинальных записей потребуется около 17 Гб.

Для совпавших записей мы производим операцию слияния. Слияние осуществляется по свойствам формата MODS. Для библиографической записи в формате MODS существуют простые и составные свойства. Для составных свойств происходит объединение наборов данных из различных записей. Для простых свойств возможны две ситуации. В случае, когда свойство присутствует в одной записи и отсутствует в другой, мы берем свойство из той записи, в которой оно присутствует. В случае присутствия простого свойства в обеих записях, мы в обогащенную запись берем свойство из более старой записи.

### **Заключение**

Для выявления дублетных библиографических записей на множестве, имеющим большую мощность, недостаточно алгоритмов, оперирующими функциями сравнения строк – содержимым полей. Если провести предварительную конвертацию строк в числа с использованием функции из семейства LSH, дополнив алгоритм поиска по получившимся значениям оптимизацией с использованием разбиения хэша на части, можно существенно сократить количество операций сравнения. Выявив таким образом претендентов на дублетные библиографические записи, достаточно сравнить их с использованием меры Жаккара на множестве биграмм, полученных из строк полей «Автор» и «Название», для установления дублетности. Полученный

алгоритм является оптимизацией известных алгоритмов поиска нечетких дубликатов для случаи большого количества исходных данных. Созданная на его основе программная система обладает высокой производительностью. Последующие исследования могут быть направлены на создание масштабируемого алгоритма, способного разделять библиографические записи на несколько потоков и параллельно их обрабатывать.

## ЛИТЕРАТУРА

1. Рубцов Д.Н. Выявление дубликатов в разнородных библиографических источниках / Д. Н. Рубцов, В. Б. Барахнин // Вестн. НГУ. Сер. Информ. технологии – 2009. – Т. 7 – № 3 – 86–93с.
2. Серебряков В.А. Проблемы семантической интеграции библиотечных данных / В. А. Серебряков, О. Н. Шорин // Библиотекосведение – 2014. – № 5 – 41–47с.
3. Charikar M.S. Similarity estimation techniques from rounding algorithms / M. S. Charikar // Proc. thirty-fourth Annu. ACM Symp. Theory Comput. - STOC '02 – 2002. – 380–388с.
4. Fellegi I.P. A Theory for Record Linkage / I. P. Fellegi, A. B. Sunter // J. Am. Stat. Assoc. – 1969. – Т. 64 – № 328 – 1183–1210с.
5. Hylton J. Identifying and merging related bibliographic records / J. Hylton // Tech. Report. Massachusetts Inst. Technol. Cambridge, MA, USA – 1996.
6. Indyk P. Approximate Nearest Neighbors: towards removing the curse of dimensionality (prelim) , 1999. – 27с.
7. Pi B. SimHash-based Effective and Efficient Detecting of Near-Duplicate Short Messages / B. Pi, S. Fu, W. Wang, S. Han // Science (80-. ). – 2009. – Т. 7 – 20–25с.

# ПОСТРОЕНИЕ ПАРАЛЛЕЛЬНОГО АЛГОРИТМА ИНДЕКСИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ FRACTAL TREES

К.В. Кузнецов, В.М. Михелев

*Белгородский государственный национальный исследовательский университет  
(НИУ «БелГУ»)*

С каждым годом размер информации, необходимой для хранения, только растет. Главной задачей связанной с хранением информации является поиск. Для реализации быстрой операции поиска необходимы дополнительные структуры данных, называемые индексами. Индексы можно разделить на индексы, использующиеся для хранения данных обычных (для сортировки которых достаточно оператора «меньше»), таких как строки, числовые типы и индексы для хранения геометрических примитивов (таких как точки и полигоны). Такое разделение охарактеризовано видом запросов поступающих к индексам. Например: запросы на входление числа в диапазон (range query) или запрос на входление в многоугольник. В этой статье рассматриваются индексы, хранящие обычные типы данных.

Во многих современных СУБД для построения индексов выбирают такую структуру данных как В-дерево. Существует много разновидностей В-деревьев, одна из наиболее используемых это В+дерева. Отличием является, то что все данные ключей содержатся только в листьях.

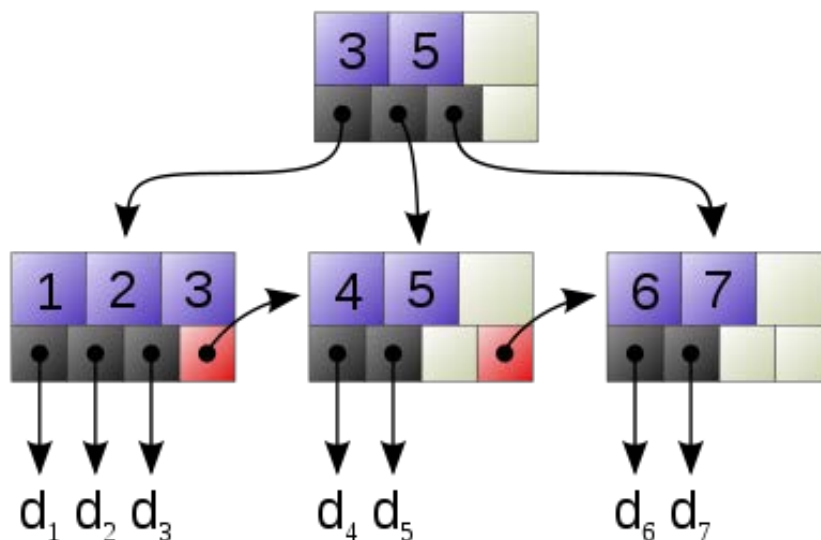


Рис. 1. Схема стандартного В+дерева.

На рисунке представлена структура В+ дерева. Два листовых узла заполнены не до конца, что отражается на скорости поиска. Также существует 2-3 дерева, в которые дается гарантия, что все узлы будут заполнены на две

третьих. Основные проблемы В-деревьев проявляются при работе с жестким диском. В этой статье будут сравниваться только структуры, хранящиеся в оперативной памяти.

Стоит отметить, что на текущее время существует множество разновидностей В-деревьев. В этой статье сравнение будет вестись с В+деревьями. В В+деревьях значение ключей хранятся только в листьях, а во внутренних узлах хранятся только ключи разделители. Это очень удобно для хранения данных на диске, так как при хранении данных на диске листовые узлы можно располагать последовательно, что существенно ускорит доступ к диску.

Другой структурой, применяющейся в СУБД, являются fractal tree index. Эта структура представляет собой набор массивов. Длины массивов равны степеням двойки. Важным ограничением при построении индекса является то, что или подмассив заполнен полностью или пуст.

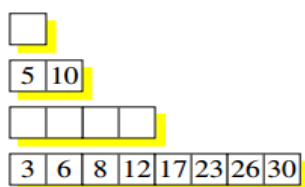


Рис. 2. Схема стандартного fractal tree index.

Такое представление данных позволяет ускорить вставку в индекс. Допустим, первый подмассив пуст, тогда вставка происходит в него, если он заполнен, то происходит слияние двух одноэлементных массивов (первый массив – подмассив структуры данных, второй – представляем вставляемый элемент как массив) и результат слияния складывается в подмассив с длиной равной двум. Существуют ситуации, при которых необходимо выполнить несколько слияний.

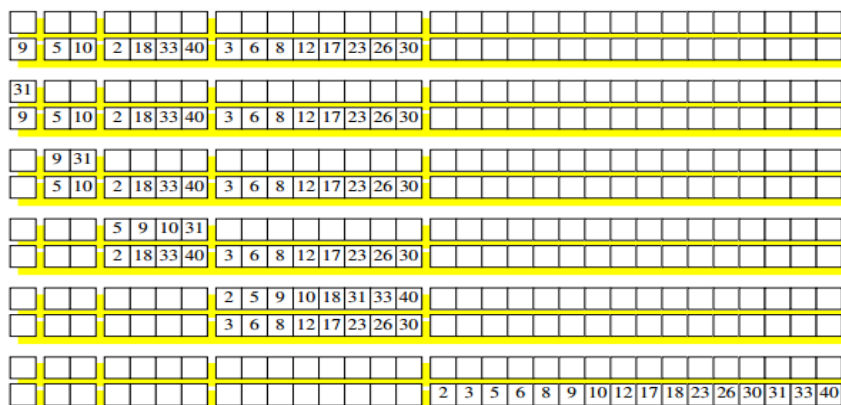


Рис. 3. Схема слияния элементов при вставке.

На рисунке хорошо видно, что выполнилось 4 слияния. Поэтому операция вставки занимает в лучшем случае 1 операцию, а в худшем –  $n$  операций, где  $n$  – текущее количество элементов в индексе. Сложность операции вставки –  $\log(N) \cdot \log(N)$ .

Предложенный алгоритм состоит в том, чтобы создавать лес из fractal tree index. Такой подход поможет легко распараллеливать операцию вставки.

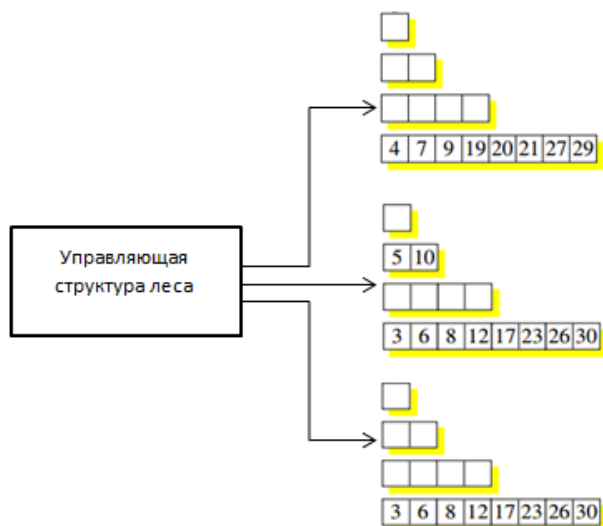


Рис. 4. Схема леса из fractal tree index

На рисунке представлено дерево из 3 fractal tree index. Лучше выбирать количество деревьев кратное степеням двойки.

Под управляющей структурой понимается алгоритм или структура данных, которая будет распределять запросы на деревья леса. В статье описан самый простой для реализации метод – равномерно распределять запросы на вставку по деревьям, запросы на удаление и выборку необходимо в такой реализации проводить ко всем деревьям.

Под параллельным алгоритмом понимается возможность параллельной вставки в индекс. Параллельность вставки достигается путем выбора дерева, не занятого вставкой. Дерево при вставке блокируется полностью, но так как количество деревьев довольно велико, то ожидание на вставку в дерево практически не происходит. Так же такой подход позволяет выбирать количество деревьев в индексе в зависимости от аппаратного обеспечения (увеличивать количество деревьев для большого количество ядер процессора).

Первый эксперимент – эксперимент на вставку. Индексирование проходило по целочисленному ключу. Размер объекта – 16 байт. Размер ключа — 16 байт. Такая ситуация позволяет при чтении индекса не сканировать таблицу (Index Only Scan). В вычислительном эксперименте участвовал лес из 16 деревьев. В этом эксперименте участвовали только последовательные реализации. Сравнивались только 3 типа индексов — B+tree, fractal tree и лес из

fractal tree. В сравнении не участвовали индексы, не поддерживающие range запросы. Так же в эксперименте не участвовали LSM и GIST индексы. Платформа для эксперимента — Intel i7 3970 (8 логических потоков), 16 Гб оперативной памяти, HDD.



Рис. 5. График зависимости времени вставки от размера индекса.

Второй эксперимент - на скорость поиска. Тип запроса – запрос на вхождение в диапазон (range query). Пример условия  $1000 < \text{число} < 2000$ . Схема выполнения запросов представлена на рисунке 6. На ней отчетливо видно что, и при запросе на равенство и при запросе по диапазону, в B-tree необходимо спустится до листовой вершины. Следующим шагом при выполнении запроса является последовательное чтение индекса. Стоит отметить, что в случае если необходимо взять из индекса 1 элемент, то в B-tree будет произведен спуск по всему дереву.

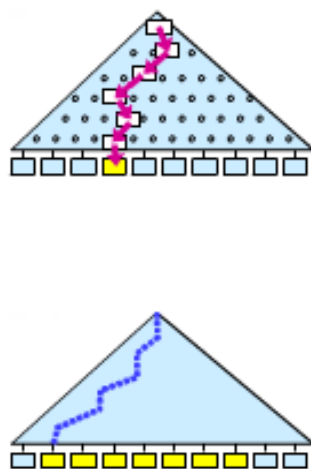


Рис. 6. Схема выполнения запросов в B-tree



При спуске в дереве внутренние узлы очень сложно хранить в последовательном порядке, что ведет к тому, что головка диска часто перемещается большими скачками, а также не происходит предсчитывание блоков, что обычно происходит при последовательном чтении данных с жесткого диска.

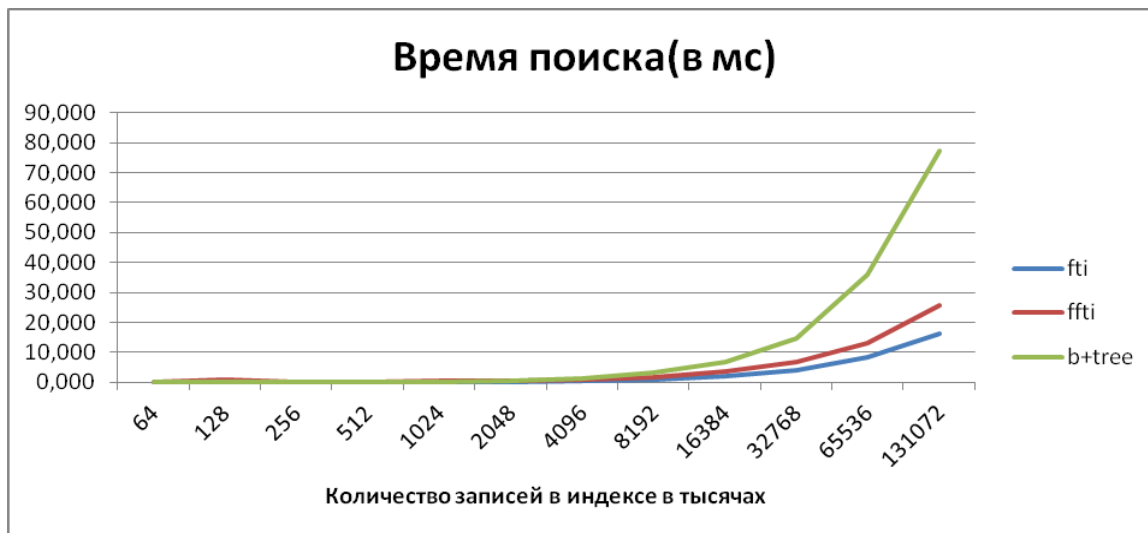


Рис. 7. График зависимости поиска от размера индекса.

В этом эксперименте участвовали только последовательные реализации. Следующий эксперимент показывает ускорение при параллельной вставке в индекс. В качестве технологии для распараллеливания выбрана OpenMP.

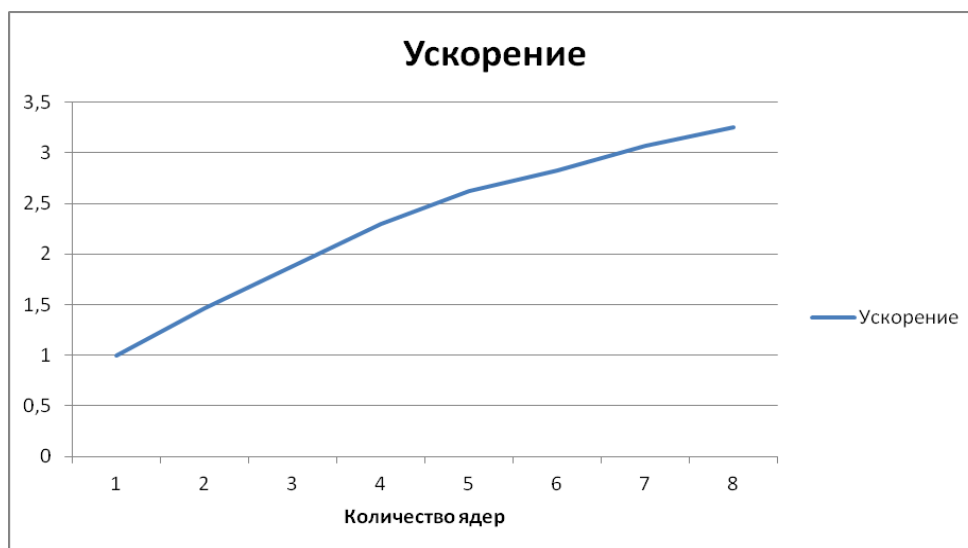


Рис. 7. График ускорения.

Количество объектов в индексе 65536 тысяч. Количество деревьев в лесу – 16. Стоит заметить, что индекс не блокируется полностью, а блокируется лишь часть – одно дерево. Как видно из графика, ускорение не

пропорционально количеству ядер, но увеличивая количество ядер, можно так же увеличивать и количество деревьев в лесу.

Выводы: разработанная структура данных имеет много преимуществ в сравнении со стандартными структурами данных, такими как B+tree, так и новыми, как fractal tree index:

- Оптимизация под жесткий диск, что повышает скорость вставки и скорость доступа.
- При хранении в оперативной памяти скорость вставки и доступа не проигрывает B-tree, но незначительно уступает стандартному fractal tree index.
- Возможность производить параллельную вставку, не нарушая целостности.
- Широкая возможность масштабирования как в системах с общей памятью, так и в системах с распределенной памятью за счет увеличения количества деревьев в лесу.

## ЛИТЕРАТУРА

1. Bender, M. A.; Farach-Colton, M.; Fineman, J.; Fogel, Y.; Kuszmaul, B.; Nelson, J. (June 2007). "[Cache-Oblivious streaming B-trees](#)". Proceedings of the 19th Annual ACM Symposium on Parallelism in Algorithms and Architectures ([CA](#): ACM Press): 81–92.
2. Esmet, J.; Bender, M.; Farach-Colton, M.; Kuszmaul, B. (June 2007). "[The TokuFS Streaming File System](#)". Proceedings of the 4th USENIX Conference on Hot Topics in Storage and File Systems ([MA](#): USENIX Association): 14–14.
3. Brodal, G.; Fagerberg, R. (Jan 2003). "[Lower Bounds for External Memory Dictionaries](#)". Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms ([N.Y.](#): ACM Press): 546–554

# ПОВЕДЕНИЕ АГЕНТОВ В ОБЛАКЕ ИНТЕРНЕТ-ОБРАЗОВАНИЯ

Г.С. Курганская

*Иркутский государственный университет,*

Облачные технологии стали уже общепринятым инструментом работы в Интернет. Главным образом это относится к организации хранения информации пользователей и, соответственно, доступа к ней, а также простых инструментов по ее обработке. На наш взгляд, наибольший – синергетический – эффект может дать интегрированное использование всех возможностей облачных технологий. Именно такой подход реализуется в интеллектуальной Интернет – платформе ГЕКАДЕМ, которая разрабатывается в иркутском государственном университете.

Архитектура Интернет – платформы представляет собой интегрированную мультиагентную систему, функционирующую в динамично меняющемся Интернет - облаке интеллектуальных ресурсов [1]. Как уже отмечалось, принципиально невозможно обеспечить общее централизованное управление ресурсами облака, поэтому возможным решением может быть самоорганизация как деятельности пользователей, так и функционирования компонентов облака интеллектуальных ресурсов. В основу реализации самоорганизующейся системы Интернет – обучения заложен мультиагентный подход. В данной статье будут рассмотрены принципы функционирования двух классов агентов платформы: владельцы ресурсов и пользователи ресурсов. Внутри классов агенты делятся по типам, и соответственно видам [1]. При этом поведение агентов будет определяться логикой, построенной на теории игр, где параметры стратегий игроков (агентов) динамически пересчитываются по сетям доверия в соответствии с байесовским подходом.

Очевидно, что в силу постоянной динамики облака мы не сможем построить сеть доверия «раз и навсегда», нам придется формировать ее всякий раз, когда агенту нужно будет принять решение. Как уже отмечалось [2], в этой ситуации мы будем следовать принципу «локальности», т.е. строиться будет фрагмент сети доверия, а точнее, дерево, корнем которого будет объект, для которого формируется оценка. Максимальное количество уровней, по которым будут рассчитываться соответствующие параметры узлов, может быть глобальным параметром платформы, в пределах которого агенты могут выбирать радиус локального окружения.

Рассмотрим подробнее механизм формирования динамической сетей доверия. Для этого сначала следует понять, какова структура облака интеллектуальных ресурсов, т.е. определить множество составляющих и отношений на них. Будем следовать KFS модели представления знаний, в которой базовым элементом является учебный блок

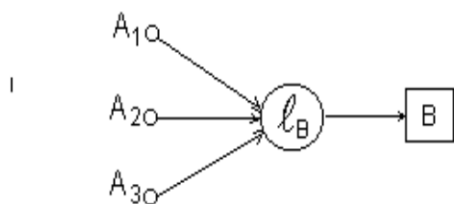


Рис.1 Учебный блок

где  $B$  - результат изучения учебного блока,  
 $A_i$  - входные знания, необходимые при изучении блока.

Блок может состоять из конечного множества других блоков, в этом случае он будет рассматриваться как кластер, в котором единственный выход и четко определены входы, если они есть. В дальнейшем мы будем рассматривать все интеллектуальные ресурсы, как такие кластеры знаний.

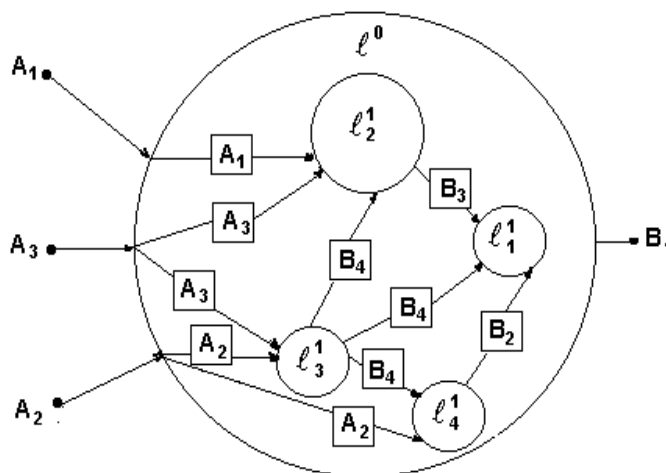


Рис.2 Кластер знаний

Таким образом, на множестве интеллектуальных ресурсов в соответствии с KFS моделью представления знаний [2] могут быть заданы два базовых отношения строгого частичного порядка:

- $part(x,y)$  – объект входит в состав объекта  $y$
- $source(x,y)$  - знания из объекта  $x$  нужны при изучении объекта  $y$ .

Очевидно, что эти отношения определяют облако интеллектуальных ресурсов как неоднородную семантическую сеть, кластеров которую можно представить для каждого отношения отдельно в виде ориентированного графа.

Отношение  $source(x,y)$  - будет определяющим для агентов – пользователей ресурсов либо при построении траектории учебного процесса,

в случае статического планирования, либо при выборе следующего кластера для изучения, в случае планирования с «колес». И в том и в другом случае алгоритм планирования по заданному графу отношения  $source(x,y)$  строит ярусно – параллельную форму (ЯПФ), предложенной Д.А.Поспеловым [3].

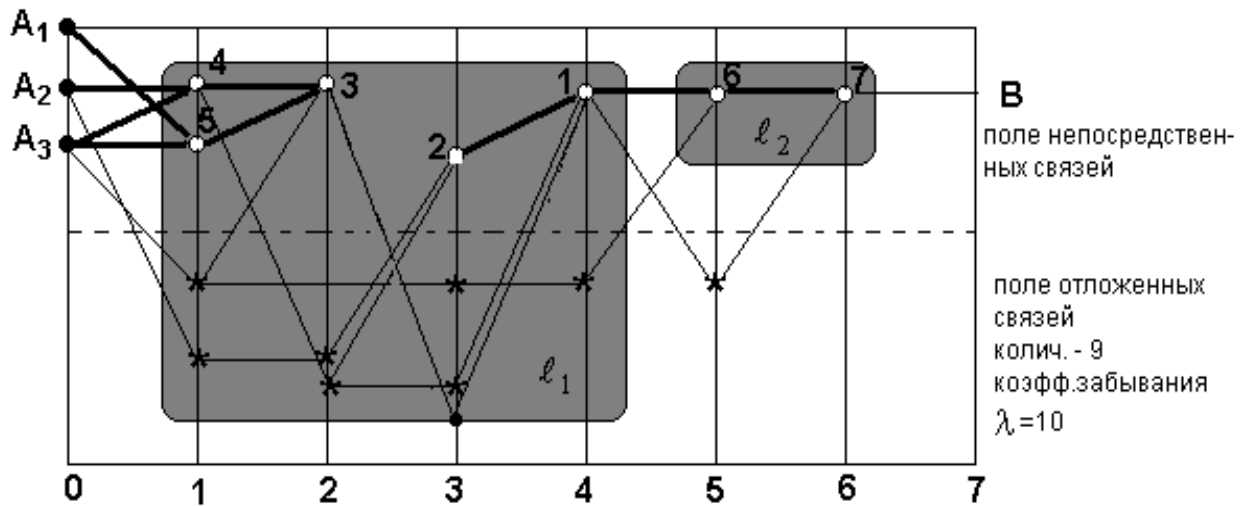


Рис.3 Ярусно-параллельная форма кластера знаний

В представленном примере видно, что почти в любой точке алгоритма есть альтернативы. Но как обеспечить рациональное поведение агентов? Каким должен быть следующий шаг? Чтобы ответить на этот вопрос, автор использует байесовский подход и теорию игр. Агенты принимают решения на основе анализа текущей ситуации, но при этом работает правило локальности, т.е для анализа доступен подграф соответствующего графа. При этом отношение  $part(x,y)$  будет служить основой для построения соответствующего фрагмента сети доверия для расчетов вероятностей условных выигрышей.

Рассмотрим взаимодействие представителей двух классов агентов, когда агенту-потребителю нужно получить ресурс. Конечно, реально у каждого игрока может быть много разных стратегий такой сделки, например:

Владелец

- продать по фиксированной цене
- продать, не дешевле чем  $min$ .
- продать подороже, но продать
- продать по любой цене
- при нескольких одновременных запросах и прочих равных условиях продать потребителю определенного типа
- .....

Потребитель

- купить самый дешевый, если предложений несколько
- купить подешевле, но надлежащего качества

- купить самое качественное, цена не важна
- купить, не дороже чем  $max$
- ...

Ситуация, очевидно, весьма неоднозначная и ее можно рассматривать в нескольких аспектах: Для простоты изложения мы не будем вдаваться в детали, хотя, конечно, именно там «кроется дьявол», а рассмотрим последовательно варианты ситуация, в которых агенты могут принимать решения, используя подходящие модели из теории игр.

### Игра с нулевой суммой

Рассмотрим ситуацию с одним ресурсом и одним претендентом. Для каждого агента определены варианты действий, и известно, что каждый может получить или потерять во всех возможных комбинациях их выбора. Интересы агентов противоположны: выигрыш одного является проигрышем другого. Это – классическая антагонистическая игра, или игра с нулевой суммой, которую можно представить в виде платежной матрицы размером  $m \times n$ , у которой каждая  $i$ -ая строка ( $i=1, 2, \dots, m$ ) отождествляется с  $i$ -ой стратегией первого игрока, а каждый  $j$ -ый столбец ( $j=1, 2, \dots, n$ ) отождествляется с  $j$ -ой стратегией второго игрока. Элементы матрицы носят смысл выигрыша первого игрока (или проигрыша второго).

Пусть платежная матрица игры имеет вид

$$H = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Понятно, что гарантированный выигрыш игрока при выборе некоторой стратегии равен минимуму соответствующей строки  $\min_j a_{ij}$ , поэтому в результате он выберет вариант с наибольшим значением такого минимума

$$\alpha = \max_i \min_j a_{ij}.$$

Величина  $\alpha$  называется нижней ценой игры – это то, что может себе гарантировать первый игрок. При максиминной стратегии.

Соответственно, второй игрок может себе гарантировать

$$\beta = \min_j \max_i a_{ij}.$$

Величина  $\beta$  называется верхней ценой игры. Соответствующая стратегия второго игрока называется минимаксной.

Понятно, что в общем случае седловую точку игры, и такое решение является оптимальным для обоих игроков.

### Смешанные стратегии

Случаи, когда имеется седловая точка, довольно редки на практике. В этом случае чаще встречаются ситуации, когда седловых точек нет, т.е. нижняя цена игры  $\alpha$  строго меньше верхней цены  $\beta$ .

Прежде отметим, что если мы обозначим через  $V$  действительную цену игры, то очевидно, что при правильных действиях игроков

$$\alpha \leq V \leq \beta$$

К такому компромиссу можно прийти, если такие сделки проводить многократно. И тогда надо менять стратегии агентов таким образом, разделить разность  $\beta - \alpha$  между двумя игроками, чтобы была максимально возможная выгода для каждого.

Процесс чередования стратегий называется **смешанной стратегией**. Каждый ход в смешанной стратегии – это выбор какой-то определенной стратегии  $A_i$  для игрока  $A$  и какой-то стратегии  $B_j$  для игрока  $B$ . Выбор стратегии – это случайный процесс, поэтому каждую выбранную стратегию на каждом шаге мы можем рассматривать как значение случайной величины вероятностью  $p_i$  для первого агента и, соответственно,  $q_j$  для второго. Понятно, что

$$p_1 + p_2 + \dots + p_m = 1.$$

$$q_1 + q_2 + \dots + q_n = 1.$$

Вектора  $P ( p_1, p_2, \dots, p_m)$ ,  $Q ( q_1, q_2, \dots, q_n)$  представляют смешанную стратегию первого и второго агента соответственно, а пара  $(P, Q)$  называется ситуацией в смешанных стратегиях. В этом случае агенты выбирают стратегии  $A_i$  и  $B_j$  независимо друг от друга случайным образом. И тогда цена игры первого агента есть математическое ожидание

$$V = \sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} P_i q_j.$$

Ясно, что цель первого агента заключается в максимизации значения  $V$ , а цель второго агента заключается его минимизации. Решение можно получить решая двойственные задачи линейного программирования и доказано, что при отсутствии седловой точки эта задача имеет единственное решение.

### **Игра с ненулевой суммой**

Если величина выигрыша первого игрока не совпадает с величиной проигрыша игрока В, то игровую ситуацию, то такой игре соответствуют уже 2 платежные матрицы, а игру называют **биматричной**. Хотя можно ее представить одной матрицей, где элементом является вектор  $(x,y)$ , где  $x$  – выигрыш первого, а  $y$  – второго игрока. Каждый агент может выбрать оптимальную с его точки зрения стратегию используя принцип максимина. В качестве оптимальных стратегий агенты могут также рассматривать те, которые соответствуют равновесию по Нэшу. Но наш взгляд, как и в случае модели антагонистической игры, лучше перейти к смешанным стратегиям, где вектора вероятностей для каждого игрока можно определить аналогичным способом.

Работа агентов по выше представленным моделям поведения опробована на макетных примерах, и получены приемлемые результаты. Понятно, что они далеко не исчерпывают все виды взаимодействий агентов платформы. Например, не рассматривали возможности формирования коалиций со своими интересами. В настоящее время ведутся исследования по привлечению более сложных моделей теории игр для решения таких задач, таких как арбитражная схема Нэша, вектор Шепли для коалиций, построение попарных соответствий (мэтчинга) по Шепли. И это позволяет автору надеяться на полную реализацию платформы облачного Интернет – обучения ГЕКАДЕМ, работающей на принципах самообучения и самоорганизации.

### **ЛИТЕРАТУРА**

1. Курганская Г. С. Архитектура мультиагентной системы интернет – обучения в облаках интеллектуальных ресурсов //Труды XIX Байкальской Всероссийской конференции «Информационные и математические технологии в науке и управлении». Часть III. – Иркутск: ИСЭМ СО РАН, 2014. – С. 205-210
2. Г.С.Курганская. Облачные технологии Интернет – образования на основе KFS модели представления знаний - Вестник Бурятского государственного университета, №9, Серия Математика и информатика – Улан-Удэ, 2013 – С. 69-76
3. Пospelов Д.А. Введение в теорию вычислительных систем. - М.: Советское радио, 1972. – 280 с.
4. Оуэн Г. Теория игр, М. Мир, 1971. – 230 с.
5. Курганская Г.С. Методы формирования оценок сложных систем по комплексному критерию в дистанционном образовании и их визуализация // Вычислительные технологии. – Том 5, №1. Новосибирск, 2000. – С. 65-70.



# СИСТЕМНАЯ ПОДДЕРЖКА РЕШЕНИЯ БИЗНЕС-ЗАДАЧ В ГЛОБАЛЬНОЙ ИНФОРМАЦИОННОЙ СЕТИ<sup>4</sup>

**Е. М. Лаврищева, Л. Е. Карпов, А. Н. Томилин**

*Е. М. Лаврищева <lavr@ispras.ru>*

*Институт системного программирования РАН,  
109004, Россия, г. Москва, ул. А. Солженицына, д. 25.*

*Московский физико-технический институт (технический университет),  
141700, Московская область, г. Долгопрудный, Институтский пер., 9.*

*Л. Е. Карпов <tak@ispras.ru>*

*Институт системного программирования РАН,  
109004, Россия, г. Москва, ул. А. Солженицына, д. 25.*

*Московский государственный университет им. М. В. Ломоносова,  
119991, Москва, ГСП-1, Ленинские горы, д. 1.*

*А. Н. Томилин <tom11@bk.ru>*

*Институт системного программирования РАН,  
109004, Россия, г. Москва, ул. А. Солженицына, д. 25.*

*Московский государственный университет им. М. В. Ломоносова,  
119991, Москва, ГСП-1, Ленинские горы, д. 1.*

**Аннотация.** Рассмотрены сервисы как системная поддержка процессов разработки программных систем и сервисы Интернет (или веб-сервисы) для решения бизнес-задач и приложений в среде Интернет. Дано описание набора системных сервисов CORBA для управления объектами и интерфейсами (stub и skeleton) между клиентом и сервером. Представлены архитектуры сервисов SOA и SCA, используемые для композиции систем из сервисов и компонентов в Интернет среде. Приведены языки описания сетевых сервисов (XML, SOAP, BPMN, WSDL, BPEL и др.) консорциума W3C и протоколы Icontract WCF для передачи сообщений между распределенными системами Интернет. Рассмотрены сервисные протоколы, и технологии обработка Web-сервисов в Grid-системе, обеспечивающей поддержку научного сервиса в глобальной сети e-sciences.

**Ключевые слова:** сервис, системный сервис, веб-сервис, модели сервисных архитектур, распределенная система; сетевая служба; протоколы взаимодействия; языки описания сервисов, бизнес задачи и приложения.

## 1. ВВЕДЕНИЕ

Системная поддержка процесса программирования задач началась с создания отдельных наборов последовательностей фрагментов программ, которые полезны всем пользователям вычислительной машины и относятся к службам управления обработкой прикладных программ, занесения их в

---

<sup>4</sup> Работа поддержана грантами Российского фонда фундаментальных исследований № 14-07-00606 и № 15-07-02355.

долговременную память, в стандартные библиотеки, а также предоставления их многим пользователям. Постепенно сформировалось понятие системного сервиса (или "службы"), которое со временем приобрело стандартное представление, содержащее информацию о реализованной системной функции, форме ее применения при решении разных вычислительных задач и бизнес приложений. Одним из наиболее известных средств обеспечения системного сервиса по управлению процессом разработки объектных программ и систем является система CORBA (<http://www.corba.org>). Она предоставляет сервисы по управлению разработкой систем на разных языках программирования, обеспечивает взаимодействие разных систем при решении некоторой задачи и передачи данных через интерфейсы IDL, обрабатываемые брокером объектных запросов ORB. Система находит применение как самостоятельно, так и внутри других общих систем [1-6].

В связи с распространением сети Интернет появилось огромное количество других видов сервисов, которыми управляют Web сервера. В e-science возник научный проект, под названием Grid (2000 г.) с целью создания компьютерной инфраструктуры нового типа, обеспечивающей глобальную интеграцию информационных и вычислительных ресурсов, специального программного обеспечения и набора стандартизованных служб для обеспечения совместного доступа к географически распределенным ресурсам. Разработан принципиально новый подход к обработке огромных объемов экспериментальных данных, к моделированию сложнейших физических процессов и созданию бизнес-приложений с большими объемами вычислений. Grid включает набор систем по реализации разных научных задач. Одной из систем является ETICS. Для построения сетевых приложений (Web Application) в ней представлены сетевые службы (Web Services), доступные глобальным пользователям [6-9].

За последние 10 лет комитет W3C разработал набор стандартов (протоколов) и языков описания программ, процессов и технологий для решения разного рода задач и бизнес-приложений в среде Интернет [10-12].

Далее рассматриваются системные и прикладные сервисы, которые представляются в рамках систем CORBA, W3C и Grid.

## **2. СИСТЕМНЫЕ СЕРВИСЫ**

Системные сервисы обеспечивают решение разного рода прикладных, деловых и бизнес-задач. К ним относятся:

- общие сервисы системных сред для поддержки процессов и функций обработки программ и данных (например, службы именования, каталогизации и др.);
- объектные сервисы, которые управляют объектами, классами и услугами по формированию и обработке объектно-ориентированных

- систем (например, службы диспетчеризации объектными запросами, управления интерфейсом и др.);
- сетевые сервисы стандартной модели OSI, моделей SOA (Service-oriented Architecture), SCA (Service-Component Architecture), как инструменты представления и обработки ресурсов в сети Интернет, которые реализуют деловые, финансовые, экономические и другие услуги при решении соответствующих задач;
  - готовые программные и информационные ресурсы (services, artifacts, reuses, assets и др.), используемые как многоразовые услуги при решении разных задач в e-science и других прикладных областях.

Некоторые из сервисов стали обязательной частью общесистемных средств (VS.Net, IBM, Intel, Linux и др.), другие используются в специальных областях (например, медицина, биология) в плане предоставления услуг при работе с современными данными (FDT, GDT, Big Data).

Каждая служба определяется именем, по которому осуществляется поиск в распределенной среде пространства имен через транзакции, устанавливающие соответствие “имя-объект” для организации и управления отдельными сервисными ресурсами глобальной сети, а также с помощью сообщений для визуального общения с требуемыми представителями отдельных ресурсов.

Перечисленные виды сервисов используются при моделировании программных систем (ПС) из готовых ресурсов сети Интернет. Для их использования при создании ПС требуется проводить поиск подходящего сервисного ресурса, его апробацию и встраивание в прикладную программу решения задачи, либо использовать его в динамическом режиме (см. [5, 6-10]).

## **2.1. Системные сервисы CORBA**

Полный набор сервисов для объектов создан в системе CORBA (см. 1-6)). В ней впервые реализована объектная модель и системные сервисы для работы с объектами:

- брокер объектных запросов (*Object Request Broker* — ORB), обеспечивающий взаимодействие объектов в разных языках программирования;
- общие объектные сервисы (*Common Object Services* — COS), обеспечивающие сервис всем объектам по вопросам управления изменениями, реализациями, контролем,
- транзакциями, подпроцессами и т.п.;
- общие средства обслуживания (*Common Facilities* — CF) или общие услуги, предоставляющие ряд общих прикладных функций, которые могут объединяться в различные конфигурации с учетом заданных требований (например, средства печати, БД, электронная почта и др.);

- объектные приложения (*Application Objects* — АО), к которым относятся приложения и их компоненты, реализующие задачи и объекты пользователя, функционирующие в объектной среде, и над которыми могут производиться операции типа - открыть, инсталлировать, переместить и поместить.

Общие объектные сервисы и общие средства обслуживания способствуют разбиению приложения на функции — базовые для большинства приложений, либо достаточно общие для широкого класса приложений. Объектные приложения и общие средства обслуживания обеспечивают функции и сервисы объектными интерфейсами, а также позволяют адаптироваться к новым поколениям сетей, языков и сред.

Система CORBA ориентирована на обеспечение взаимодействия удаленных объектов в распределенной среде типа OpenStep. Для задания взаимодействия объектов используется язык интерфейсов IDL (*Interface Definition Languages*) объектов. Интерфейсы в IDL запоминаются в репозитории интерфейсов (*Interface Repository*), а реализации объектов — в репозитории реализаций (*Implementation Repository*). Независимость интерфейсов от реализаций объектов позволяет использовать их разными приложениями статически и динамически.

Объект-клиент и объект-сервер обмениваются между собой с помощью запросов, каждый из которых исполняется брокером ORB с помощью компонентов, создаваемых на основе описания интерфейсов клиента, сервера и ядра ORB.

**Интерфейс клиента** (*Client Interface*) обеспечивает взаимодействие с объектом-сервером и состоит из трех базовых интерфейсов:

- stub-интерфейса, содержащего описание внешне видимых параметров и операций объекта в IDL;
- интерфейс динамического вызова (*Dynamic Invocation Interface* — ДИ) объекта, определяемого во время выполнения программы клиента посредством поиска описания интерфейса в репозитории интерфейсов или в репозитории реализаций;
- интерфейса сервисов ORB (*ORB Services Interface*), содержащего набор сервисных функций, которые клиент запрашивает у сервера через брокер ORB.

**Stub-интерфейс** обеспечивает взаимосвязь клиента с ORB. Прикладная программа клиента через посредника (заместителя) *stub* — статической части программы клиента, посылает в запросе параметры, которым сопоставляются соответствующие описания из репозитория интерфейсов. Брокер ORB выполняет полученный запрос и пересылает результаты клиенту.

Рассмотренная схема взаимодействия клиента с объектом соответствует схеме вызова удаленных процедур через RPC-механизм.

**Интерфейс ДП** обеспечивает доступ объектов и их интерфейсов во время выполнения. Этот интерфейс предоставляет механизм запроса к объектам, интерфейс которых становится известным во время выполнения. Он становится доступным с помощью вызовов брокера ORB. В каждом вызове указывается тип объекта, тип запроса и параметры. Эту информацию посылает прикладная программа, но она может извлекаться из репозитория интерфейсов или репозитория реализаций.

Объекты специфицируются средствами языков программирования (см. [1, 4, 5]) и могут быть реализованы на разных платформах и в разных средах. Интерфейсы программ-посредников описываются на языке IDL. Заместитель клиента (*stub*) выполняет сервисные функции, связанные с преобразованием типов данных клиентских компонентов к стандартным системным типам, а заместитель сервера (*skeleton*) преобразует стандартное представление данных в типы данных сервера.

Описание интерфейса в IDL начинается с ключевого слова **interface**, за которым следует: имя интерфейса, описание типов параметров и операций (*op\_dcl*) вызова объектов:

```
interface A { ... }  
interface B { ... }  
interface C: B, A { ... }.
```

Параметры операций (*op\_dcl*) в задании интерфейсов это:

- 1) тип данных (*type\_dcl*);
- 2) константа (*const\_dcl*);
- 3) название исключительной ситуация (*except\_dcl*), которая может возникнуть в процессе выполнения метода объекта;
- 4) атрибуты параметров (*attr\_dcl*).

Описание типов данных (ТД) начинается ключевым словом **typedef**, за которым следует базовый или конструируемый тип и его идентификатор. В качестве константы может быть некоторое значение типа данного или выражение, составленное из констант. ТД и константы описываются как фундаментальные типы данных: *integer*, *boolean*, *string*, *float*, *char* и др.

Описание операций *op\_dcl* передачи данных включает в себя:

- 1) наименование операции интерфейса;
- 2) список параметров (от нуля и более);
- 3) типы аргументов и результатов, иначе – *void*;
- 4) управляющий параметр или описание исключительной ситуации и др.

Атрибуты передаваемых параметров начинаются служебными словами: **in** – при отсылке параметра от клиента к серверу; **out** – при отправке параметров-результатов от сервера к клиенту; **inout** – при передаче параметров в оба направления (от клиента к серверу и обратно).

Описание интерфейса для одного объекта может наследоваться другим объектом и тогда это описание становится базовым, например:

```
const long l=2
interface A { void f (in float s [1]); }
interface B { const long l=3; }
interface C: B, A { }.
```

Интерфейс **C** использует интерфейс **B** и **A**. Это означает, что интерфейс **C** наследует описание типов данных **A** и **B**, которые по отношению к **C** являются внешними. При этом синтаксис и семантика остаются неизменными. Согласно приведенному примеру - операция функции  $f()$  в интерфейсе **C** наследуется из **A**.

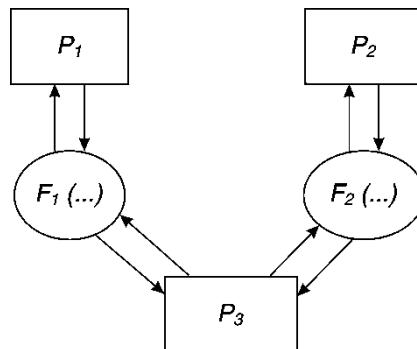
Механизм наследования интерфейса состоит в сохранении имен объектов без их переопределения. Это касается описания операций, которые должны иметь уникальные обозначения. Имена операций могут использоваться динамически во время выполнения интерфейса *Skeleton*.

Общая структура описания модуля с интерфейсом в языке IDL имеет вид:

```
Request Operations
module CORBA {
    interface Request {
        Status add-arg (
            in Identifier name,
            in Flags arg_flags
        );
        Status invoke (
            in Flags invoke_flags // invocation flags
        );
        Status send (
            Status get_response (
                out Flags response_flags // response flags
            );
        );
    };
};
```

Тип данных описывается в классе FDT, GDT, которые передаются через параметры операторов RPC, RMI, а также протоколами в WCF VS.Net, либо еще какими-нибудь. ТД описывается на объектно-ориентированном языке программирования (C#, VBasic, Pascal, и др.).

Входные и выходные интерфейсы (например) для программ  $P_1$ ,  $P_2$  (рис.1) имеют разную семантику, но одинаковое синтаксическое описание на некотором языке программирования. Передача данных от этих программ для  $P_3$  осуществляется через функции  $F_1 (...)$ ,  $F_2 (...)$  и интерфейсы *In*, *Out*, с помощью которых осуществляется преобразование ТД переданных между  $P_1$ ,  $P_3$  и  $P_2$ ,  $P_3$  туда и обратно.



**Рис. 1.** Схема вызова функции объектов

Данный аппарат интерфейса реализован в системах CORBA, ОС IBM, Microsoft и др. Его основу составляют библиотеки VS.Net (CRL, CTS, FCL, CIL и др.) преобразования ТД, которые применяется при интеграции разнородных программных объектов.

Общие объектные службы предоставляют набор операций для работы с разными категориями объектных приложений:

- наименование и поиск объектов по именам, либо по свойствам и атрибутам;
- управление жизненным циклом объектов;
- параллельного обращения к объектам;
- определение очередей запросов к объектам;
- иерархия транзакций;
- защиту объектов от несанкционированного доступа (авторизация и аутентификация клиентов и др.).

Интерфейсные средства IDL используются для обеспечения взаимодействия распределенных систем прикладного, коммерческого и бизнес типа, а также входят в состав общесистемных систем (VS.Net, Java и др.),

### **3. СЕТЕВЫЕ ТЕХНОЛОГИИ ИНТЕРНЕТ. СТАНДАРТЫ W3C**

Сеть Интернет базируется на стеке протоколов TCP/IP, за которые отвечает международная некоммерческая организация ISOC (Internet Society).

ISOC насчитывает более чем 20 тысяч представителей, свыше 100 организаций из 180 стран мира и предоставляет организационную основу для других консультативных и исследовательских групп в том числе:

IETF открытое международное сообщество ученых, инженеров, провайдеров которое занимается развитием протоколов и архитектуры Интернет;

ICANN (Corporation for Assigned Names and Numbers) — международная некоммерческая организация, которая координирует предоставление имен и адресов в Интернете.

Структура TCP/IP базируется на четырехуровневой модели сетевого взаимодействия, которое разработано Министерством обороны США и в основном отвечает 7 уровневой модели OSI.

В Интернете используется общий язык для передачи данных глобальной информационной среды, стандартные методики и форматы представления данных и обмен протоколами. На рис.2 приведены организации, генерирующие сетевые Интернет технологии.

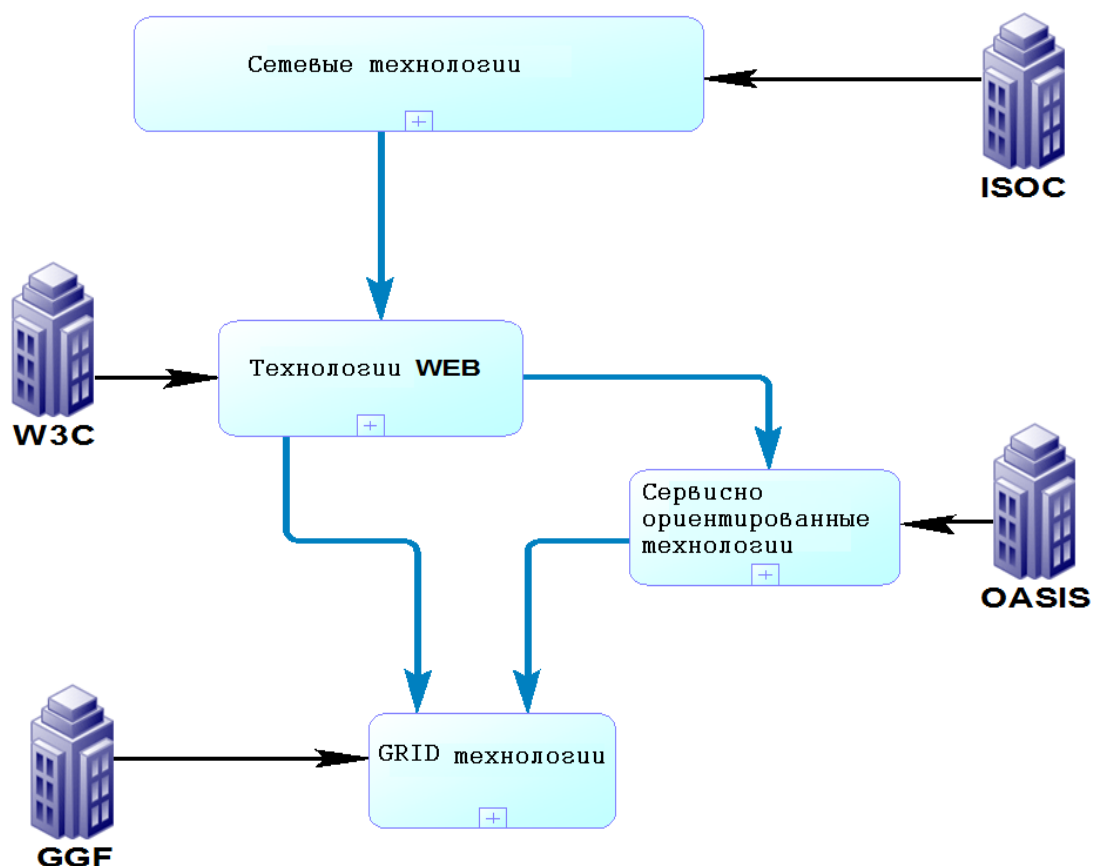


Рис.2. Организация формирования стандартных Интернет технологий



### 3.1. WEB технологии

Глобальная информационная сеть (WEB-среда) объединяет десятки миллионов документов по всему миру и развивается благодаря стандартам консорциум W3C [12-20]. Консорциум W3C объединяет наиболее крупных производителей программного обеспечения для WEB технологий (Web серверов и WEB браузеров). W3C обеспечивает компьютерным программам взаимодействие в сети (так называемая «сетевая интероперабельность»).

Стандарты глобальной информационной сети можно подразделить на 4 группы.

#### 1). Представление форматных данных пользователя

HTML язык разметки гипертекста. и документов. в виде ASCII текста, фрагменты которого облагаются специальными пометками (тегами);

XHTML расширяемый язык разметки гипертекста (Extensible Hypertext Markup Language) на основе принципов и синтаксиса XML;

CSS каскадные таблицы стилей (Cascading Style Sheets) для писания внешнего вида документа (страницы);

MathML (Mathematical Markup Language) – язык математических формул, который использует формат XML для отображения математических формул. Используется для отображения формул в Web браузерах MathML;

SVG (Scalable Vector Graphics) – язык масштабируемой двумерной векторной графики, которая использует формат XML.

#### 2). Представление структурированных данных

XML Schema – язык для определения правил, которым должен удовлетворять документ и состоять из набора тегов и их атрибутов для описания соответствия документа его предметной отрасли.;

XLink (XML Linking Language) – XML схема, которая рекомендована W3C для организации ссылок между ресурсами;

XInclude (XML Inclusions) – XML схема, которая рекомендована W3C и предоставляет механизм включения в XML-документы текстовых файлов и др.;

XSL (eXtensible Stylesheet Language) – семейство рекомендаций (стандартов) которое определяет методологию превращения XML документов для визуализации или другой обработки соответствующими программными средствами;

- XSLT декларативный язык, который интерпретирует правила и применяет их к входным документам;
- DOM (Document Object Model) – не зависящая от платформы и языка программирования модель, который позволяет создавать динамические Веб страницы посредством скриптовых языков (JScript, JavaScript);
- XML Encryption – определяет порядок шифровки и дешифровки содержания элементов XML документа;.
- XKMS (XML Key Management Specification) – определяет порядок безопасной передачи и регистрации открытых ключей для шифровки и дешифровки XML документов за протоколом XML Encryption;.
- PNG (portable network graphics) – формат для сохранения растровой графики;
- SMIL (Synchronized Multimedia Integration Language) и язык мультимедийной интеграции на синтаксической основе формата XML..

### **3). Протоколы удаленного выполнения программ (сервисов)**

- SOAP (Simple Object Access Protocol) [15, 16] – определяет порядок обмена сообщениями (данными) в WEB среде.;
- WSDL (Web Services Description Language) [16] – язык для описания Веб сервисов и доступа к ним на основе XML.

### **4). Представление семантических данных**

- RDF (Resource Description Framework) – ряд стандартов, который определяет базовые методы формального представления знаний для машинной обработки;
- OWL (Web Ontology Language) – язык, который определяет правила описания онтологии предметных отраслей и предназначен для обеспечения одинаковой и однозначной интерпретации документов разными агентами в распределенной среде.

Наиболее распространенным является протокол SOAP. При описании сервиса указывается адрес URI (Uniform Resource Identifier) и транспортный протокол (например, HTTP). Средством описания функциональности сервиса является язык WSDL (Web-service description language, см. [15]). Для представления данных, в особенности метаданных, используется модель RDF (см. [16]). Описание процессов представления и обработки запросов на сервисы в графическом виде осуществляется языками:

WSCCI (Web Services Choreography Interface,

WSCL (Web Services Conversation Language,  
BPMN (Business process and model and notation,  
BPEL (Business Process Execution Language for Web Services, см. [20-24]) и др.

### 3.2. Сервисно-ориентированная архитектура (SOA) и SCA

*Модель SOA* (Service-oriented Architecture) задает сервисно-ориентированную архитектуру программной системы, а *модель SCA* (Service-Component Architecture) – архитектуру на основе сервисов и компонентов. К средствам моделирования сложных систем относится система WebSphere Integration Developer компании IBM. Она предоставляет сервисно-ориентированную архитектуру SOA и SCA, в виде use case языка UML Эта система обеспечивает интеграцию сервисов SCA через модель интерфейсов JAVA, задаваемую в языке WSDL и классах JAVA. Эта модель дает доступ к сервисным компонентам и определяет зависимость между ними через аппарат ссылок. Они упаковываются в модуль для выполнения на сервисном модуле WebSphere Process Server, который эквивалентен EAR-файлу J2EE и некоторым другим. Подмодули J2EE и артефакты упаковываются с модулем SCA, который затем запускает сервис и передает данные для их интеграции [24].

Механизмы, которые используются для вызова внешнего сервиса, названного *импортом* и *экспортом*, связаны с другими технологиями, такими как JMS, Enterprise JAVA Beans или веб-сервисы. SCA модуль может обратиться к существующему Enterprise JAVA Bean для обеспечения релевантного представления в *универсальной модели данных*, а также обмена данными друг с другом через SDO. В этой модели объекты данных представлены в JAVA common.sdo.DataObject и включают в себя метод, который позволяет пользователям получать свойства данных. WebSphere Integration Developer используется на платформе Eclipse 3.0.

#### 3.2.1. Модель SOA

*Модель SOA* предоставляет набор принципов и средств создания системного ПО и прикладных ПС из совместимых и унифицированных сервисов.

Модель группируется на серверной стороне из некоторого количества согласованно реализованных сервисов и их служб. Группы задают открытый интерфейс, содержащий описание типов входных/выходных параметров каждого сервиса и портов обмена метаданными в языке WSDL. Для WSDL созданы компиляторы, позволяющие получать серверные и клиентские заместители и учитывающие особенности конкретных программных платформ, в том числе языки программирования на этих платформах, которые описывают

реализуемые операции.

Унификация сервисов состоит в типизации функциональности сервиса и его характеристик, а также языков описания сервисов и их взаимодействия. Объект SOA обладает специфицированной функциональностью и качеством. В его модели используются две технологии, которые обеспечивают функциональность (Functions) и качество сервисов (Quality service). Эти технологии вынесены на уровень IT-стандартов комитета W3C.

Технология обеспечения *функциональности* веб-сервисов включают в себя:

- 1) транспортный уровень (transport layer) для обмена данными;
- 2) коммуникационный уровень (service communication layer) протоколов;
- 3) сервисный уровень (service description layer) и связанные с ним интерфейсы;
- 4) уровень бизнес-процессов (business process layer) для реализации бизнес-процессов и потоков работ через механизмы веб-сервисов;
- 5) уровень реестра сервисов (service registry layer), который обеспечивает библиотеку веб-сервисов для их публикации, поиска и вызова WSDL-описаний интерфейсов.

Технология обеспечения *качества* веб-сервисов имеет следующие уровни:

- 6) политики (policy layer) для описания правил и условий применения веб-сервисов;
- 7) безопасности (security layer) для описания вопросов безопасности веб-сервисов и функционирования (авторизация, аутентификация и распределение доступа);
- 8) транзакций (transaction layer) для установления параметров обращения к веб-сервисам и обеспечению надежности их функционирования;
- 9) управление (management layer) веб-сервисами.

Технологическую основу составляют: XML, SOAP, UDDI, WSDL. С их помощью осуществляется реализация базовых свойств веб-сервиса и механизмов взаимодействия между собой веб-сервисов в среде SOA (рис.3).

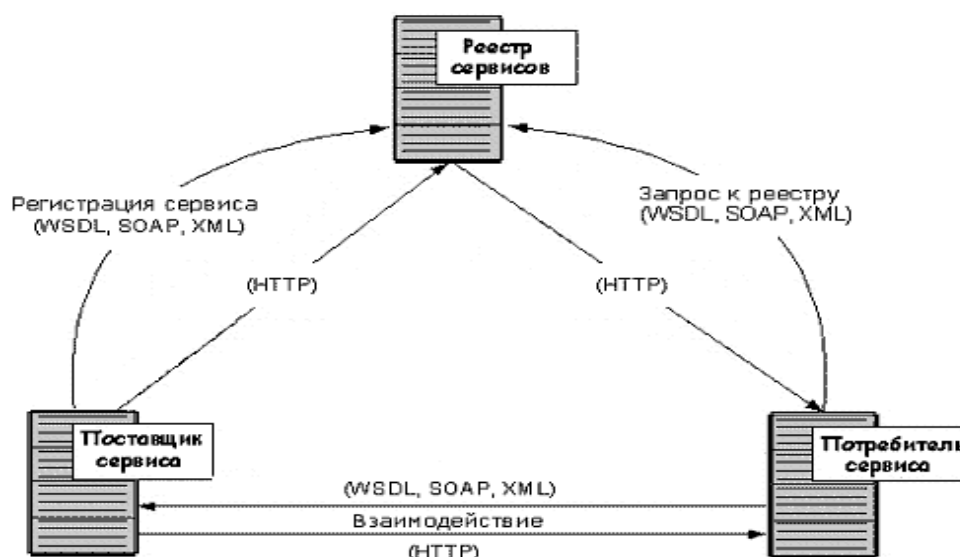


Рис. 3. Поставщики и потребители сервисов в сети

К ним относятся:

- 1) провайдер сервиса осуществляет реализацию сервиса, прием и выполнение запросов пользователей, публикацию сервиса, от из реестра сервисов;
- 2) реестр сервисов содержит библиотеку сервисов для поиска и вызова сервиса по запросам от поставщика или провайдера сервисов, предоставляющих сервисы;
- 3) потребитель или пользователь сервиса осуществляет поиск и вызов необходимого сервиса из реестра описания сервисов, а также использует сервис, предоставленный поставщиком в соответствии с его интерфейсом (рис. 3).

Посредником между этими службами и приложениям является *провайдер*, который обеспечивает взаимодействие между поставщиками и провайдерами с помощью средств описания и передачи сервисов WSDL, SOAP, XML.

Для получения сервиса в архитектуре SOA выполняются следующие операции:

- публикация сервиса WSDL с целью обеспечения доступности пользователю сервиса и его интерфейса;
- поиск сервиса в реестре с помощью протокола SOAP и заданных критериев;
- связь с реестром UDDI через описание пользователем необходимого сервиса.

При этом предусматривается, что в реестре архитектуры SOA содержится описание сервиса с форматом запросов пользователя к провайдеру, содержащему в себе перечень описаний сервисов, которые могут быть вызваны соответственно с опубликованным интерфейсом сервиса.

К базовым функциям управления компонентами и службами в операционной среде относятся:

- поиск необходимых ресурсов (компонентов, reuses, assets, artifacts, служб и др.);
- доступ к названным ресурсам;
- организация обмена информации между компонентами, ресурсами и службами;
- динамическое управление функционированием заданной совокупности ресурсов.

Модель сервисов ПС базируется на унификации и совместимости, что позволяет рассматривать ПС как набор сервисов, функциональности и взаимодействия. Унификация достигается путем:

- типизации функциональности сервиса и их других характеристик;
- применения унифицированных языков для описания сервиса и их взаимодействия;
- использования стандартных базовых технологий.

Отдельный класс средств унификации составляет онтология – модели и словари, которые обеспечивают согласование терминов и понятий языка описания сервисов на уровне семантики. В качестве стандартных базовых технологий при реализации сервисов используются: модель клиент-сервер, унифицированные коммуникационные протоколы, компонентные модели и т. д.

Смысл SOA состоит в том, чтобы создавать небольшие компоненты и собирать их в большой распределенный по глобальной сети комплекс под бизнес-задачи конкретного клиента. Функциональные сервисы, как обычные программные модули, могут быть распределены по вычислительным системам и обладать способностью к взаимодействию посредством локальных и/или глобальных сетей. Интерфейс таких модулей не зависит от технологии или платформы, в рамках которой они реализованы.

Бизнес-процесс определяется как набор взаимосвязанных задач, относящихся к деловой деятельности, имеющий начальные и конечные точки для повторения задачи.

### 3.2.2. Сервисно-компонентная архитектура (SCA)

Модель SCA предназначена для работы с компонентами со спецификациями, разработанными различными компаниями: компонентами EJB сервера приложений J2EE, сетевыми сервисами, объектами планирования, доступа к базам данных, к информационной системе предприятия и др. [2, 24].

Архитектура SCA обеспечивает доступ к сервисным компонентам и определяет зависимости между ними через аппарат ссылок. Механизмы, которые используются для вызова внешнего сервиса, называются импортом и экспортом. Элементы SCA могут компоноваться и обмениваться данными друг с другом, пересылая сервисные объекты данных, подготовленные в необходимом виде. Этот интерфейс включает определение метода получения и установления свойства данных. В рамках модели SCA сервисы могут собираться в различные образования (хореографии). Они используют архитектуру SOA и/или создают новые сервисы для их комбинирования и конфигурирования (рис.4).

Характеристики приложения можно согласовывать со специфическими потребностями клиента (конфигурирование), используя такие элементы, как стандарты бизнеса, бизнес-правила, соглашения о бизнес-сервисе и параметры конфигурации (рис. 2).

*Модель SCM* представляет собой обобщение объектно-компонентной модели семейства программных продуктов (см. [1]). В ней каждый член является: системой удаленных *peuses*, которые обмениваются гетерогенными данными и предоставляют сервисы реализации с множеством общих свойств; композицией сетевых сервисов, которые поддерживают некоторый деловой процесс. Данная модель ориентирована на обеспечение адаптивности ПС к переменным условиям использования запросов к функциям и обрабатываемым ими гетерогенных данных. В интересах этой поддержки для программной реализации ПС используются механизмы сервисных объектов данных и сервисов доступа к ним. Они позволяют размежевать код ПС и код доступа к обрабатываемым данным. В состав модели SCM входят подмодель *абстрактных сервисов*, *интерфейсная* подмодель и подмодель *объектных ресурсов*.

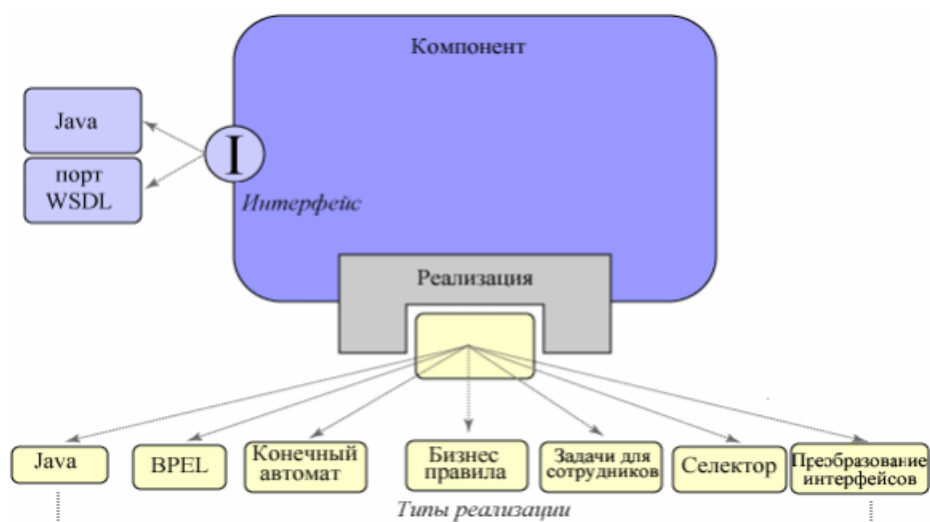


Рис. 4. Общая схема сервера SCA IBM

Сервер приложений Java Enterprise Edition содержит набор спецификаций на языке Java, которые необходимы при работе с сетевыми программами и средствами. К ним относятся:

- динамическая генерация серверных страниц (Java Server Pages);
- сетевые службы;
- компоненты повторного использования (Enterprise Java Beans);
- служба обмена сообщениями (Java Message Queue) и другие сервисные технологии.

Сетевая служба идентифицируется с помощью универсального ресурсного идентификатора URI, ее ресурсы (свойства и методы) описываются на специальном языке WSDL. Доступ к ресурсам осуществляется через протокол SOAP, который представляет собой XML-запросы, передаваемые посредством интернет-протоколов относительно высокого уровня (HTTP, SMTP). Сетевые службы соответствуют объектам объектно-ориентированных языков программирования с некоторыми важными отличиями.

Ключевым понятием сетевой службы **SCA** является сообщение, которое состоит из одной или нескольких переменных. Вместо методов классов в сетевых сервисах используются операции, которые определяются входным и выходным сообщениями. Для описания общедоступных ресурсов сетевых сервисов в язык WSDL, построенный на синтаксической основе языка разметки XML, введены возможности описания данных различных типов. В качестве переменных для сообщений можно использовать последовательности, созданные из фиксированного количества переменных простых типов, причем типы, которые будут использоваться в службе, декларируются заранее.



### 3.2.3. Среда WCF для взаимодействия систем

Программная среда Windows Communication Foundation (WCF) входит в состав среды .NET Framework (см. [25]) и является логическим развитием технологий сетевых служб, .NET Remoting и DCOM. В основе WCF лежит модель SOA, которая обеспечивает на сервере работу некоторого количества сервисов, определенных в интерфейсе для задания абстрактных входных/исходных параметров. Эти операции описываются на WSDL и могут быть сделаны доступными через, так называемые mex-endpoints (Metadata Exchange Endpoints), что позволяет получить "метаданные" сервиса. Подключаясь к этому интерфейсу; можно получить описание сервиса и всех его операций, а также сгенерировать соответствующий прокси-класс (класс-заместитель) для заданного языка или платформы. Сервис описывается в языке WCF и используется с языками Java / Python / Ruby и т. п. Клиенты в свою очередь имеют на своей стороне прокси-классы, которые содержат ссылку на операции к сервису.

В WCF MS.Net работает технология фабрик программ (AppFab), основанная на нескольких веб-службах и фабрике сервисов. Основу технологии составляют схемы, "рецепты", методы и средства построения программ разного назначения. Фабрика программ включает в себя набор ресурсов, блоков кода, документации, образцов приложений, инструментов и паттернов VSIP для создания из них разных пакетов, которые накапливаются в глобальном хранилище Global Bank. Фабрика сервисов содержит рекомендации, схемы и методы, а также стандарты по проектированию и конструированию продуктов. Функционирование WCF определяется так называемыми конечными точками (endpoint), которые устанавливают "ABC" или "Address-Binding-Contract" (см. рис.5).

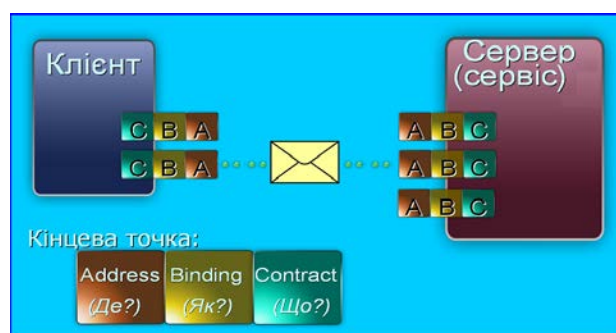


Рис.5 Точки в WCF

Каждая точка выполняет определенную роль:

"Address" задает место расположения конечной точки по абсолютному или относительному адресу;

"Binding" задает связь с транспортным протоколом, на основе которого будет происходить взаимодействие. Связь в виде классов-привязок (BasicHttpBinding в HTTP, NetTcpBinding в TCP и т. д.);

"Contract" задает контракт, на основе которого будет происходить взаимодействие клиента и сервиса с помощью операций сервиса, который строит класс-прокси на стороне клиента.

На сервисной стороне задается множество конечных точек. В результате, сервисная сторона распределенного приложения будет выглядеть как совокупность конечных точек.

Инструменты распределенного действия WCF или .Net Remoting построены по принцип «слоенного пирога», каждый слой которого отвечает за свой конкретный уровень абстракции и не знает ничего о нижележащих уровнях. Инфраструктура WCF состоит из двух главных уровней: Service Model Layer и Channel Layer. Первый уровень расположен ближе к самому сервису клиента и обеспечивает превращение метода с параметрами в сообщение для передачи более низкому уровню. Канальный уровень (Channel Layer) инкапсулирует канал передачи данных, которых может быть много, такие как транспорт TCP, Http, Named Pipes и т. д. Каждый из этих уровней содержит подуровни, и может включиться в каждый из них.

Контракты представляют собой описание сообщений, переданных конечными службами с возвратом. Конечная точка должна специфицироваться и может выполняться в формате ожидаемых данных. Совокупность этих спецификаций и есть контракт.

WCF содержат три вида контрактов:

- 1) контракты сервисов для описания функциональных операций, реализованных сервисом. Внутри контракта имеются операции сервиса, которые реализуют функции;
- 2) контракты данных определяют формат данных, которыми будут обмениваться сервисы. Это относится к запросу на сервис и октету сервиса. Если используются примитивные типы – *int*, *string* и др., то контракт не нужен, потому что в .Net имеется возможность сериализации и десериализации типов. Для комплексных типов – *Customers*, *Order* и др., необходимо указать принцип сериализации и десериализации;
- 3) контракты сообщений, как тип контракта, который используется для того, чтобы получить контроль над заголовком SOAP (рис.6).

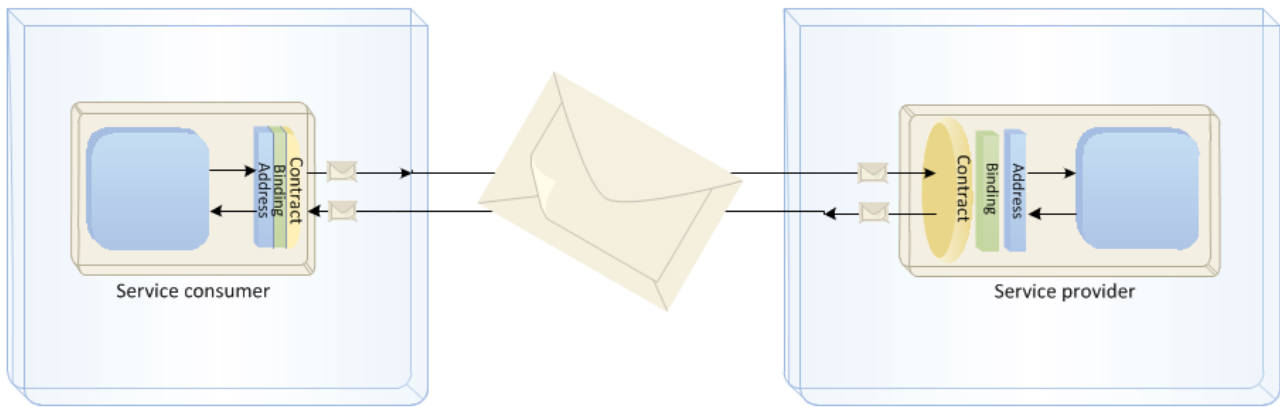


Рис. 6 Передача пакета сообщения между потребителем и поставщиком

В пакете передается протокол SOAP в виде:

```

<?xml version="1.0" ?>
<env:Envelope xmlns:env="http://www.cbsystematics.com">
<!--Конверт протокола SOAP-->
<env:Header>
<!-- Заголовок протокола SOAP-->
</env:Header> <env:Body>
<!--Тело протокола SOAP-->
</env:Body> </env:Envelope>.

```

При взаимодействии систем общего назначения друг с другом через сообщения в конверте могут возникнуть разного конфликты передачи данных (рис. 7).

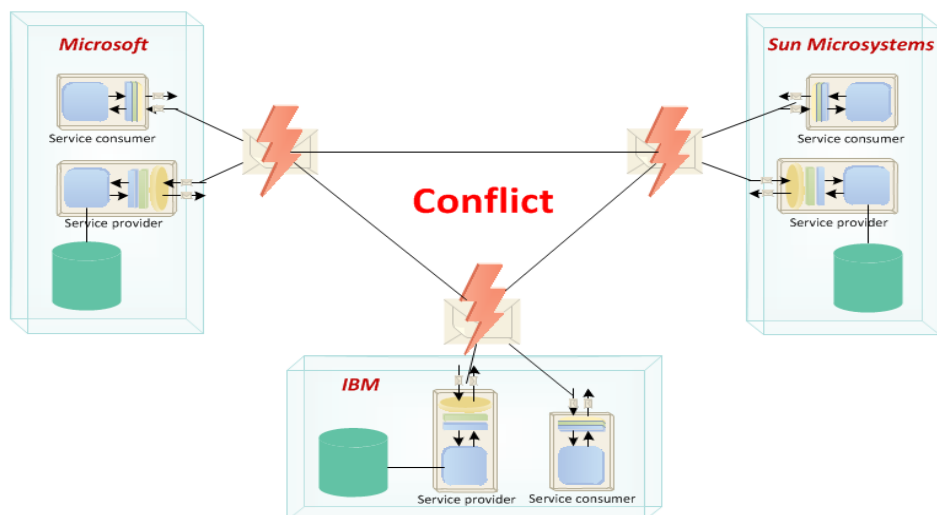


Рис.7. Схема взаимодействия систем с возникновением конфликтов

Например:

- 1) Несовместность типов данных, которые передаются в контрактных интерфейсах потребителя (вместо целого присылается символьный).
- 2) несоответствие в порядке (или количестве) параметров в протоколе передачи;
- 3) Разлиха в архитектуре платформ объектов клиента и сервера.

Для обеспечения интероперабельности контрактов в широком диапазоне систем, используются языки WSDL и XSD. При написании кода службы поставляется класс с определенными в WCF атрибутами [ServiceContract], [OperationContract], [FaultContract], [MessageContract] и [DataContract] и утилита *svcutil.exe*, которая вызывает конечную точку сервиса для возврата данных после генерации WSDL-документа по атрибутам.

На этапе выполнения вызывается метод, определенный в интерфейсе сервиса, WCF сериализует типы CLR и вызов метода в формат XML и посылает сообщение в сеть для привязки к схеме кодировки, согласованной с WSDL. При этом участвуют четыре конструкции: две со стороны .NET и две со стороны XML. В .NET имеется тип CLR, который определяет структуры данных и функциональные возможности создания объект такого типа. Со стороны XML задается XSD-описание структуры данных, но сообщение осуществляется лишь после того, как будет создан экземпляр XML (XML Instance).

**Язык WSDL.** Для описания общедоступных веб-ресурсов используется язык WSDL (Web Service Definition Language) на основе XML. Среда программирования Eclipse позволяет автоматически создавать описания на основе классов Java. В языке определены следующие основные типы данных:

- 1) строки (xsd:string);
- 2) целые числа (xsd:int, xsd:long, xsd:short, xsd:integer, xsd:decimal), числа с плавающей запятой (xsd:float, xsd:double);
- 3) логический тип (xsd:boolean);
- 4) последовательности байт (xsd:base64Binary, xsd:hexBinary);
- 5) дата и время (xsd:time, xsd:date, xsd:g);
- 6) объекты (xsd:anySimpleType).

В качестве переменных для сообщений могут использоваться последовательности, созданные из фиксированного количества переменных простых типов.

Типичный WSDL-файл имеет следующую структуру.

```
<wsdl:definitions [.]>  
<!-- Декларация типов, которые используются в сервисе -->  
<wsdl:types>  
<element name="someMethod">
```

```

<complexType>
  <sequence>
    <element name="arg0" type="xsd:double"/>
    <element name="arg1" type="xsd:boolean"/>
  </sequence>
</complexType>
</element>
<element name="someMethodResponse">
  <complexType>
</wsdl:types> ...

```

Приведенное WSDL-описание определяет веб-сервис MyService с единственным методом String someMethod(double arg0, boolean arg1). На его основе можно сгенерировать два типа данных, которые отвечают входным и исходным аргументам метода. Эти типы применяются в описаниях someMethodRequest и someMethodResponse – входного и выходного сообщений для операции someMethod.

Операции декларируются в описании интерфейса сервиса (декларация wsdl:portType) и дальше в описании привязки сервиса к SOAP (декларация wsdl:binding), причем во втором случае также оговаривается способ вызова (<wsdlsoap:body use="literal"/>). За счет этого при вызове операции используются те же названия параметров, что и в методе класса. В конце WSDL-файла находится декларация веб-сервиса (<wsdl:service>), в которой содержится информация относительно его расположения (параметр location).

### 3.3. Технологии Grid

Документ Open Grid Services Architecture (OGSA) [2, 8, 9], определяет концептуальную основу системы Grid и ее основных компонентов (рис.8): архитектуру, данные, инфраструктуру, компьютеры и приложения. Система Grid объединяет локальные ресурсы, которые настраиваются и администрируются автономно (нижний слой), объединяются, руководствуются и динамически отслеживаются за счет программного обеспечения промежуточного слоя (middleware).

Прикладное программное обеспечение через стандартизованные интерфейсы имеет доступ к локальным и распределенным ресурсам через средний слой, который скрывает от пользователя внутреннюю сложность и физическое отображение ресурсов. В основу стандарта OGF положена архитектура SOA, согласно которой вся функциональность промежуточного слоя реализуется путем комбинации взаимоувязанных Web сервисов (SOAP, WSDL). Служба OGSA (Basic Execution Service) определяет интерфейс к сервисам, которые иницируют вычислительные процессы в Grid, отслеживают и руководят вычислительной деятельностью. Кроме того определены модели

жизненного цикла (состояния) процессов, а также информационные модели вычислительного процесса.

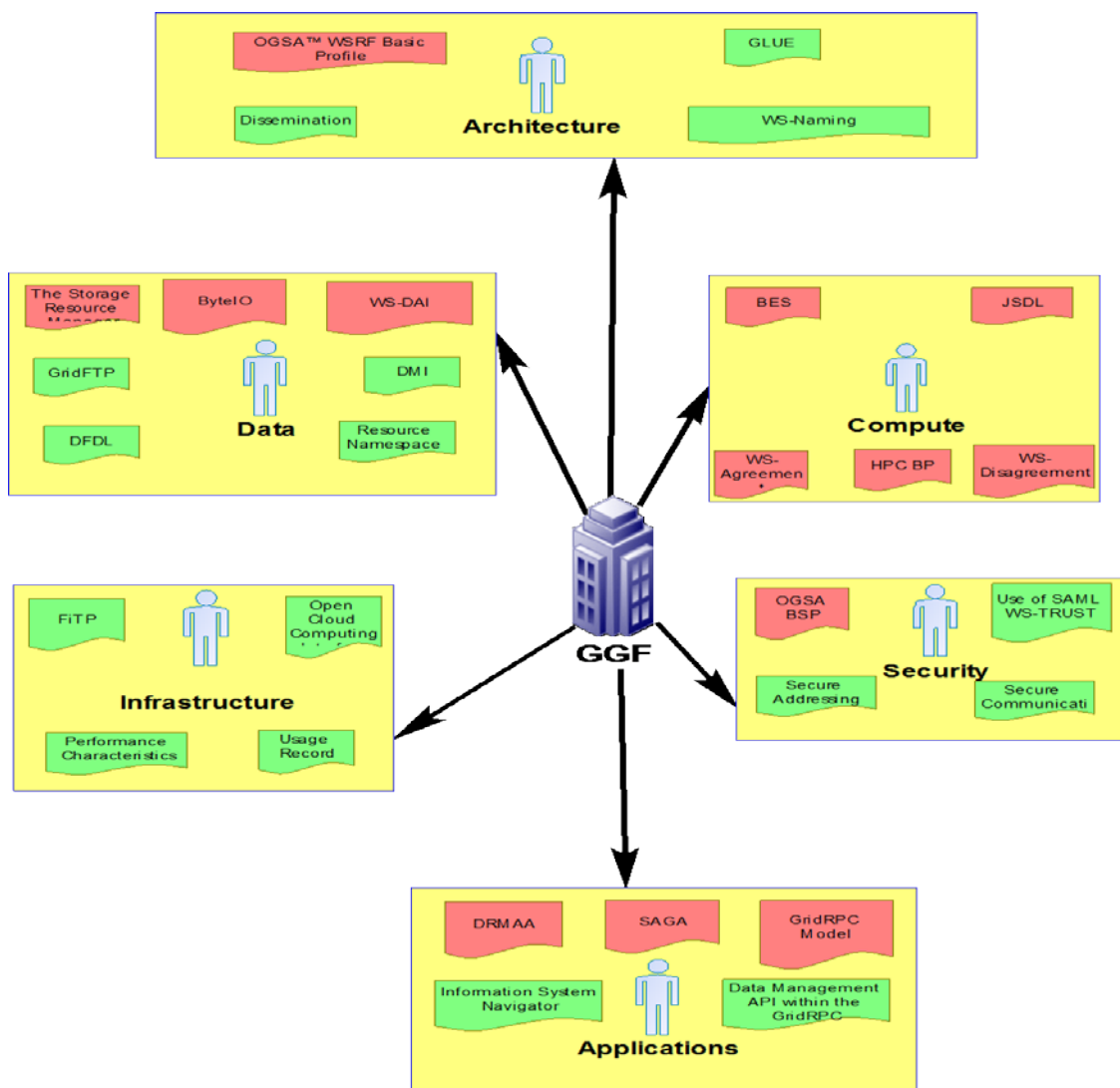


Рис. 8. Структура стандартов Grid

В Grid реализован программный интерфейс SAGA (Simple API for Grid Applications). Он определяет программную модель для описания компонентов Grid и сигнатуру методов для поддержки процесса выполнения заданий, управления ресурсами и др. Целью спецификации является предоставление стандартизованного доступа к функциям Grid на языках высокого уровня (C, Java, Python и т.п.) и тем самым обеспечение их совместимости с разными middleware.

Одной из Grid систем является ETICS. Она предоставляет инфраструктуру (Web Application и Web Services) для тестирования, интеграции и конфигурирования, занесения в репозиторий готовых программ и ПС (рис. 9). ETICS включает процессы сборки (builds) и управления тестами на компьютерах пользователей, обеспечивают отправку удаленных сборок на

разные платформы пользователей [26]. Программные объекты (пакеты, компоненты и системы) разрабатываются по модели CIM (Common Information Model).

Сценарии работы пакетов и систем задаются на языке моделирования UML. Сервисный набор процедур Etics обеспечивает построение и тестирование новых пакетов и систем с помощью механизмов плагинов. Каждый плагин включает в себя публичный интерфейс и описание услуг.



Рис. 9 Портал ETICS

В набор характеристик системы входят механизмы спецификации зависимостей между разными пакетами и их тестами. Функциональные плагины обеспечивают проверку договоров, тестов для выполнения разных элементов систем, генерацию документации и ведения готовых объектов в оперативном или постоянном репозитории. В ETICS разработаны наборы пакетов из комбинаций перекомпилированных двоичных элементов, размещенных в репозитории. Компоненты регистрируются в репозитории в стандартном языке WSDL консорциума W3C (рис.10). Стандартное описание компонентов используется при проведении сборки и тестирования (Build and Test Web Application) групп компонентов. При конфигурации совокупности компонентов в результирующий файл поступают команды контроля версий (Command-Line Client), тестирования свойств и зависимостей в выходном коде системы. Этот сервис можно задать командной строкой задания сборки и тестирования (Build and Test Service). Удаленная сборка/тестирование выполняется на разных платформах с генерацией ряда отчетов о сборке,

статическом и динамическом тестировании отдельных компонентов в разных средах глобальных пользователей.

Каждый новый компонент содержит сведения (имя, лицензия, URL репозитория), глобальный уникальный идентификатор – ID (GUID); информацию о версиях, GUID (свойства, среде выполнения и зависимости), тестовых команд и GUIDs.

Типы данных пакетов и систем объединяется плагинами, в которых задаются услуги для потребителей или поставщиков, доступ к ОСАМ, архитектуре CPU, компиляторам с языков программирования и средствам спецификации зависимостей между разными пакетами, используемым при сборке программ и их развертывании. Плагины обеспечивают проверку контрактов, тестов выполнения разных элементов систем, генерацию документации, ведения готовых КПИ в репозиториях.

Главная проблема в ETICS – преобразование некоторых компонентов систем для альтернативной платформы гетерогенной среды компьютеров путем ссылок с 16-, 32-разрядной платформы на 64-разрядную платформу среды Grid.

**Request a new external component** << Back Next >>

**Step 2/4 - new component**

Please fill following form providing new component details.  
Press **Next** button when the form is ready (the button is enabled **only** when the form is properly filled).

**Information about the new component**

*name:*   
Name to uniquely identify the module inside the project.  
This field is mandatory. Have to be a combination of letters, numbers, \_ , - , . (dot).

*display name:*   
More descriptive name (optional).

*description:*   
A free form description. This field is mandatory.

*vendor:*   
Vendor name. This field is mandatory.

*license type:*   
Type of license - for instance *Apache License 2.0*. This field is mandatory.

*homepage:*   
URL to module homepage (for information purposes only).

*download:*   
URL where the module can be downloaded (for information purposes only).

<< Back Next >>

Рис. 10. Стандарт описания нового компонента в Grid

#### 4. ЗАКЛЮЧЕНИЕ

В работе рассмотрены сервисы, которые являются системной поддержкой процессов разработки программных систем и веб-сервисы Интернет как информационные ресурсы, используемые для решения научных и бизнес-задач в среде Интернет. Дано описание набора системных сервисов CORBA для управления объектами при обмене данными (через *stub* и *skeleton*) между



клиентом и сервером. Представлены модели сервисов SOA и SCA, используемые для композирования систем из сервисов и последующего их выполнения в распределенной среде. Дано описание системы Grid, в состав которого входит ряд систем поддержки физического эксперимента и реализации прикладных систем в e-sciences (физики, биологии, математики, медицины и др.). В ней реализованы операции сборки (building), тестирования (testing) и конфигурации (configurating) и работа с данными, согласно универсальной модели данных. Приведено описание ряда стандартов W3C, включая протоколы и языки описания сервисов для функционирования в среде Интернет (XML, SOAP, BPMN, WSDL, BPEL и др.). Приведено описание протоколов (Contract) в системе WCF для передачи сообщений между распределенными системами и обеспечения взаимодействия между разнородными системами в среде Интернет.

## ЛИТЕРАТУРА

1. Андон Ф.И. Методы инженерии компьютерных распределенных приложений.-К: Наук.думка.-1997. 271 с.
2. Лаврищева Е. М. Software Engineering компьютерных систем. Парадигмы, Технологии, CASE-средства программирования. – К.: Наук. Думка, 2014 – 284 с.
3. Лаврищева Е. М., Грищенко В.. Сборочное программирование. Основы индустрии программных систем.- Наук.Думка, К.: 2009.-370с.
4. Карпов Л. Е. Архитектура распределенных систем программного обеспечения – М., МАКС Пресс, 2007. – 130 с.
5. Карпов Л. Е., Юдин В. Н. Обмен данными в распределённой системе поддержки решений. Труды Института системного программирования, т. 19, М., Институт системного программирования РАН, 2010, стр. 71-80, ISBN 978-0-543-57630-9, ISBN 978-5-4221-0085-9, ISSN 2220-6426 (Online), ISSN 2079-8156 (Print), [http://www.ispras.ru/ru/proceedings/docs/2010/19/isp\\_19\\_2010\\_71.pdf](http://www.ispras.ru/ru/proceedings/docs/2010/19/isp_19_2010_71.pdf)
6. Jon Siegel. "Quick CORBA™ 3". Wiley Computer Publishing, John Wiley & Sons, Inc., 2001 (Джон Сигел, "CORBA 3", М., МАЛИП, 2002).
7. Gustavo Alonso, Fabio Casati, Harumi Kuno, Vijay Machiraju. Web Services. Concepts, Architectures and Applications. Springer-Verlag, 2004.
8. А.П. Демичев, В.А. Ильин, А.П. Крюков, Введение в грид-технологии. Препринт НИИЯФ МГУ - 2007 - 11/832 2007, <http://www.sinp.msu.ru>
9. А.М. Ходжибаев\*, Ф.Т.Адылова, Новейшие информационные ГРИД-технологии в электронной медицине, Донецк, журнал телемедицины и телемедицины том 3, №1, 2005
10. Lim Gray. A transformed scientific method. [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_jim\\_gray\\_transcript.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf)

11. Дрёмин И.М. Физика на большом адронном коллайдере. УФН. Июнь 2009. т. 179 № 6 Fusion for energy. Annual report 2011. av. At [http://fusionforenergy.europa.eu/downloads/mediacorner/publications/reports/ANNUAL\\_2011.pdf](http://fusionforenergy.europa.eu/downloads/mediacorner/publications/reports/ANNUAL_2011.pdf)
12. <http://www.w3.org/TR/2008/REC-xml-20081126/>
13. Гладцын В. А., Кринкин К. В. Яновский В. В. Сервис-ориентированная архитектура: стандарты, алгоритмы, протоколы – Санкт-Петербург: СПб ГЭТУ ЛЭТИ, 2006 – 108 с.
14. Papazoglou M. P., Dubray J.-J. A Survey of Web Service Technologies, Technical Report DIT-04-058, Ingegneria e Scienza dell'Informazione, University of Trento, 2004.
15. <http://www.w3.org/TR/soap/>
16. <http://www.w3.org/TR/soap12-part1/wsdl20/RDF/wsci>
17. <http://www.omg.org/spec/BPMN>
18. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>
19. <http://www.oasis-open.org/specs/index.php#wsbpelv2.0>
20. <http://uddi.org/pubs/uddi-v3.0.2-20041019.htm>
21. <http://www.oasis-open.org/specs/index.php#uddiv2>
22. <http://www.oasis-open.org/specs/index.php#uddiv3>
23. <http://www.ibm.com/developerworks/websphere/techjournal>
24. Джим Амсден, Моделирование SOA: часть 1. Идентификация сервисов, IBM® Rational® Application
25. <http://127.0.0.1:4000/ICContract>
26. <https://web.infn.it/etics-support/index.php/documentation>

# СООТНОШЕНИЕ ИНФОРМАЦИОННЫХ И ЭКОНОМИКО-МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Ю.П. Липунцов <sup>1</sup>, В.А. Серебряков <sup>2</sup>

*1 Экономический факультет МГУ им. М.В. Ломоносова*

*2 ВЦ им. А.А. Дородницына РАН*

## 1. Введение

Ценность информационной системы определяется двумя составляющими: информационной наполненностью и интеллектуальностью обработки данных. Характеристики данных: количество, их качество определяется информационной моделью. Интеллектуальность системы определяют применяемые методы обработки данных.

В традиционных экономических системах, ориентированных на решение локальных задач, информационные и математические модели соединены в коде приложения. С ростом масштаба систем эти две категории моделей разделяются: информационная модель в большей степени ориентирована на интеграцию данных, получаемых от внешних источников, а экономико-математическая модель должна содержать аналитические инструменты для обработки данных не только локальными компонентами, но и сервисами внешних поставщиков.

## 2. Модели информационной системы

Информационные системы создаются для сбора, хранения и обработки данных. Методы обработки данных определяются на этапе математического моделирования. Концепция математического моделирования как "интеллектуального ядра" информатизации представлено в работах Самарского А.А. (Самарский, и др., 2001). В других работах по математическому моделированию приведено деление формальных моделей на математические и информационные (Трусов, 2004, стр. 32). Рассмотрим модели экономических информационных систем как совокупность двух составляющих: информационная и экономико-математическая модель.

В работах по изучению информации (Колин, 2014) реальный мир описывается как совокупность двух составляющих: материи и информации. При воспроизведении реального мира в информационной системе также создается две категории моделей: информационная модель отражает информацию реального мира, а математическая больше относится к описанию материальных объектов и их поведения.

Развитие вычислительной техники позволило совершенствовать обработку данных. Становление математического моделирования произошло в период развития космической, ядерной тематики. В условиях информационного

общества основным объектом моделирования становятся социальные - экономические явления, совокупность отношений, возникающих в ходе хозяйственной деятельности.

Экономические системы относятся к классу сложно формализуемых объектов, в которых данные выступают не только объектом обработки, но и основным критерием адекватности модели. Такие характеристики экономической системы как постоянная изменчивость, адаптация, рефлексия требуют непрерывного совершенствования системы и информационной модели. Математическое моделирование предлагает представление, которое требует постоянной проверки на реальных данных. Сочетание вероятностных методов с аналитическими сыграло существенную роль в развитии технологических систем.

Современные информационные системы, соединяющие разные категории участников, является сложными системами. В системной инженерии базовой является V диаграмма, которая представляет проект в виде трех стадий: декомпозиция, реализация и интеграция. Декомпозиция включает описательную модель, дизайн высокого уровня, детальный дизайн. Далее следует стадия реализации компонент, и затем интеграция компонент в модули, а модулей в систему. Рассмотрим с этих позиций информационную и экономико-математическую модели.

### 3. Информационное моделирование

Основным элементом информационной модели является модель отражения предметной области в слое данных. Информационную модель рассмотрим как совокупность собственной модели предметной области и модели интеграции данных других систем.

В моделировании предметной области обычно выделяют три уровня абстракции информационных моделей: концептуальное, логическое и физическое моделирование.

**Концептуальная информационная** модель предметной области, как правило, является описательной моделью, составленной посредством терминов и понятий проблемной области.

В качестве вариантов представления **логических моделей** рассмотрим модели бизнес-архитектуры, коммуникативных актов, документов, сервисов и бизнес процессов.

В основе структурного представления деятельности в формате *бизнес-архитектуры* лежит управленческое мышление, включающее формальные и неформальные правила организации. Элементами такой модели могут быть организационный и культурный контекст деятельности.

Организация как совокупность *коммуникаций* описана в онтологической модели предприятия Ян Дитца (Dietz, 2006). Организация рассматривается как совокупность координационной и операционной деятельности, значительная часть которых реализуется посредством коммуникативных актов.

Коммуникации являются объектом исследования рабочей группы ВМІ ОМГ и других работах.

Парадигмой ряда проектов является представление деятельности организации как совокупности *документов* и их потоков. В документах фиксируется значительная часть деятельности предприятия: они являются носителями корпоративных знаний, описывают управленческую деятельность, определяют технологические процессы предприятия, включают руководства для информационных систем и т.д. В ряде проектов документ рассматривается как инструмент реализации логики деятельности.

Более формальный взгляд рассматривает организацию как совокупность *сервисов* для внешних пользователей и внутри предприятия. Сервисная модель воспринимает поставщика сервиса как «черный ящик», для которого описываются функции и внешнее поведение.

Наиболее детально логику деятельности отражает идеология *бизнес процессов*. На уровне бизнес-процессов каждая компонента рассматривается как «белый ящик», содержание которого представлено детально на уровне отдельных элементов.

Совокупность информационных моделей логического уровня обеспечивают конкретно ориентированный, но платформу-независимый взгляд на информацию с позиций логических структур данных. Схемы логического уровня на поздних стадиях проектирования воплощаются в логическую модель данных.

**Физическая модель** представляет содержание логической модели в формате базы данных. Основным вариантом **реализации** является создание реляционной модели. В реляционных моделях объекты, транзакции и прочие элементы реального мира отражаются в сущностях, и детализируются в атрибутах. Дополнительными элементами модели данных являются связи и их характеристики, которые определяют типы пересечений множеств элементов сущностей.

С развитием всемирной паутины WWW разработаны технологии описания ресурсов в формате *RDF (Resource Description Framework)*, который позволяет организовать хранение больших объемов данных в универсальном формате. С использованием этой модели данных создаются такие классы приложений как википедии, социальные сети и прочие масштабные решения в среде веб.

Новой реальностью в области данных становится *Большие данные* (Big Data), возникающие вследствие индустриализации поставки данных от традиционных поставщиков данных, а также от умных устройств, снабженных датчиками, сенсорами и интерфейсами для коммуникаций.

Современные приложения предполагают **интеграцию** на уровне данных.

*Интеграция в рамках предприятия* реализуется по модели мастер данных, данных о базовых информационных объектах. На основе мастер

данных организуется взаимодействие между системами, а также управление последовательностью выполнения сервисов.

В *полу-контролируемой* среде взаимодействие может быть разным по уровню интеллектуальности. OMG (OMG, 2011) рассматривает создание *технологической инфраструктуры*, использование *единых форматов данных* и *семантические методы*.

Обмен данными посредством технологической инфраструктуры предполагает создание сервисов для обмена. В этом варианте возможно подключение новых систем, ограничением является возрастание сложности в случае роста количества сервисов. Этот тип модели реализован в СМЭВ 2.0. (Минкомсвязь РФ)

Слой данных имеет определяющее значение при реализации метода *форматов данных*. В моделях этого класса выполняется стандартизация метаданных. Эти модели обмена реализованы в National Information Exchange Model (NIXM), Statistical Data and Metadata Exchange (SDMX) (Statistical Working Group) и других проектах.

*Семантическая модель* предполагает передачу смысла при отражении реального мира в его символическое описание. На уровне данных семантика обеспечивается двумя механизмами: присвоения имен и однозначная идентификация объектов. Модели этого класса используют стандартизированные словари и единые идентификаторы для распределенной среды. На семантические методы опираются архитектурные модели обмена государственными данными Евросоюза (European Interoperability Reference Architecture) (EIRA), стандарт UPDM (OMG UPDM).

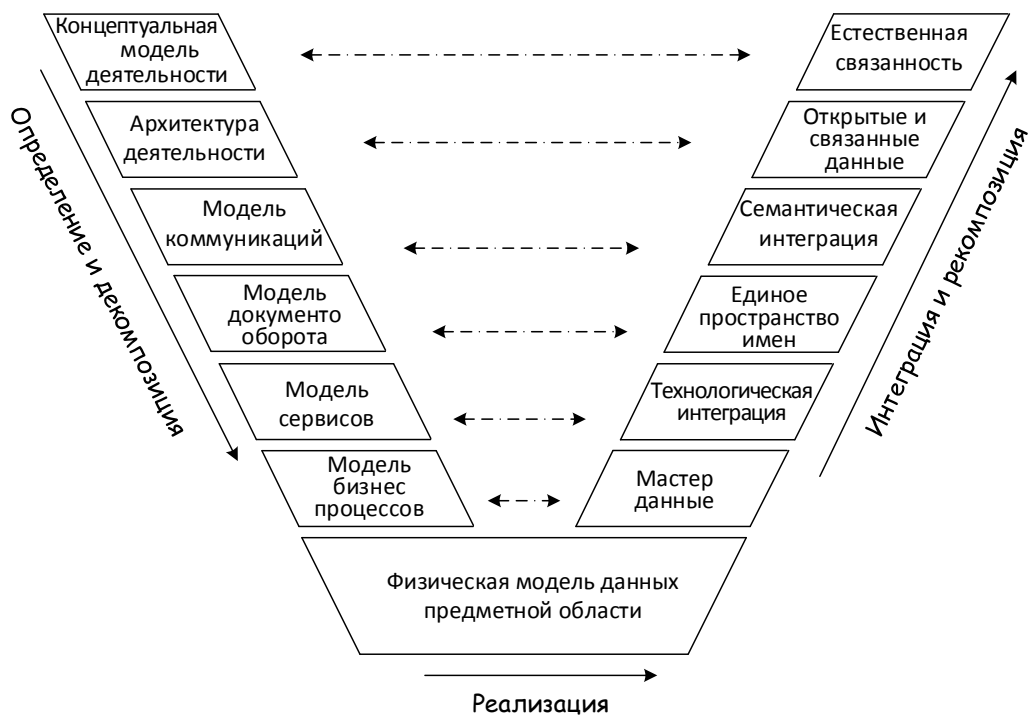


Рисунок 6. V модель информационного моделирования

Вариантом повышения информационного наполнения являются *открытые связанные данные*. Публикация накопленных данных в машиночитаемом формате позволяет повторно использовать эти данные, создавать приложения на открытых данных. Приложения на открытых данных используются словари и спецификации консорциума W3C, а механизмом связывания данных выступают открытые реестры, такие как Opencorporates, Legal Entity Identifier и другие.

V диаграмма информационной модели приведена на рис.1

#### 4. Экономико-математическое моделирование

Для выделения стадии разработки экономико-математической модели воспользуемся общей теорией систем (Месарович, и др., 1978), которая предполагает восемь стадий: лингвистическая, теоретико-множественная абстрактно-алгебраическая, топологическая, логико-математическая, теоретико-информационная, динамическая и эвристическая. Первые пять уровней будут относиться к **декомпозиции**, теоретико-информационная к **реализации**, а последние две к **интеграции**.

На верхнем уровне определяется каркас предметной области и ее границы. Модель *лингвистического* уровня описывает систему как совокупность объектов исследования и отношений между ними. Макроэкономическая модель этого уровня включает такие объекты, как государство, фирма, домохозяйство, рынок товаров и услуг, финансовые рынки и т.д., а отношения между ними описываются понятиями доход, налоги, потребление, сбережения, и т.д. Аналогичные модели строятся и для микроуровня (Грачева, и др., 2013).

На *теоретико-множественном* уровне абстракции описываются формализация элементов отдельной системы, их иерархия, определяется факт наличия связи между элементами, но не описывается ее тип. Примером из макроэкономики является совокупность индикаторов: валовый внутренний продукт, чистый внутренний продукт, национальный доход, и т.д.; на уровне предприятия: доход – затраты - прибыль – распределение и т.д.

На *абстрактно-алгебраическом* уровне связи элементов описываются посредством функций, параметров и прочих артефактов. Путем обследования реального объекта выявляются его существенные характеристики, строится содержательная модель, а на ее основе создается математическая модель. В условиях, когда информационная система представляет собой совокупность распределенных ресурсов, на этом уровне абстракции закладываются основы для интеграции отдельных вычислительных компонент для обработки данных полученных от компонент уровнем ниже.

*Топологические* модели рассматривают расположение экономических субъектов в пространстве. Примерами топологических задач могут быть модели «линейного города» и «кругового города», т.е. модели пространственной дифференциации рынка «на линии» и «на окружности».

Детальное описание функций и обработка данных с использованием математического аппарата рассматривается на *логико-математическом уровне*. На этом уровне для моделируемого объекта создается математический объект. В соответствии с экономическим явлениям ставится формальный математический объект, описанный на языке математики, а его исследование проводится с использованием тех или иных математических методов. Математические модели разделяются на аналитические и имитационные модели.

Аналитические модели представляются в виде совокупности алгебраических, интегральных, дифференциальных и иных уравнений, которые позволяют найти решение в явном виде. Возможности по описанию экономической действительности посредством уравнений ограничены, поскольку введение дополнительных параметров при соблюдении непротиворечивости математической логики вызывает трудности. Посредством уравнений удается описать статическую структуру, фундаментальные характеристики моделируемого объекта, обладающие высокой общностью.

Наиболее приближенное отражение действительности позволяют воспроизвести имитационные модели, в которых описывается последовательность действий процесса. Модель позволяет провести многократные испытания и оценить на этой основе параметры процесса. Сложные явления моделируются как совокупность локальных моделей.

Математической основой имитационных моделей является метод Монте-Карло, общая теория систем, теория массового обслуживания и др. Математический аппарат аналитических моделей включает нахождение решений для уравнений. Методы нахождения оптимальных решений используют симплексный метод, методы теории графов и оптимизации на сетях и т.д. Как инструмент оценки параметров моделей микро- и макроэкономики используется эконометрика.

Разработанные экономико-математические модели в форме алгоритмического описания готовы к **реализации** в виде программных компонент. Соединение экономико-математических моделей с информационными моделями выполняется на этапе *теоретико-информационных моделей*. Методы для организации взаимодействия вычислительных компонент с поставщиками данных представлены в следующем разделе.



Методы **интеграции компонент**, реализующих экономико-математическую обработку данных, предопределяется степенью контроля среды системы. В рамках контролируемой среды возможно использование разнообразных методов обработки данных, могут применяться как имитационные, так и аналитические методы. Для моделирования в полуконтролируемой и неконтролируемой среде в большей степени подходят методы имитационного моделирования. Аналитические методы, применяемые в период плановой экономики, требуют адаптации и соответствующего информационного обеспечения.

Принципы интеграции для предприятия, холдинга, мезоуровня и макроуровня будут разные, но рассмотрение масштаба модели не является предметом этой статьи, хотя является неотъемлемой частью экономико-математических моделей.

В контролируемой среде, на уровне предприятия, интеграция начинается с компонент оперативного учета и планирования производства, управления запасами. Более высокого уровня задачи – прогнозирование и оценка производственных мощностей, задачи бухгалтерии, управление кадрами, и прочие. Расширением этой идеологии является включение в производственный процесс поставщиков сырья и потребителей готовой продукции. Математический аппарат для обработки данных более широкого круга клиентов предполагает разработку рекомендательных систем и интернет-сервисов, использование методов кластеризации клиентов, деревья решений, анализ формальных понятий и пр. С использованием этого аппарата выполняется скоринговые оценки клиентов для управления рисками, формируются целевые предложения для клиентов при реализации маркетинговых стратегий, выполняется оптимизация колл-центров, прогнозирование остатков в разделе оперативного управления.

Развитием моделей внутренней деятельности предприятия являются имитационные модели. Имитационные модели воспроизводят *динамику*. Различают три типа имитационных моделей: событийные, агентские модели и модели системной динамики.

Событийные модели реализуют дискретно-событийный подход, рассматривающий деятельность как дискретную во времени последовательность событий. Развитием этой модели является моделирование по действиям, выполняемым по расписанию. Данные для имитационных моделей могут поставляться на основе эконометрических оценок. Другим вариантом поставки данных для имитационных моделей могут выступать информационные системы предприятий.

Более сложным вариантом имитационных моделей является *агентское моделирование*, которое рассматривает действия автономных, децентрализованных агентов. Поведение агентов моделируется с привлечением методов теории игр, теории сложных систем и их эмерджентности, математической социологии, и ряда других методов.

Более высокий уровень интеграции систем реализуется методами *системной динамики* и используется в основном для задач стратегического уровня. Системная динамика является методологией и методикой математического моделирования для построения моделей верхнего уровня, которые позволяют обсуждать сложные вопросы и понимать их.

Вариантами аналитических моделей этого класса могут выступать модели межотраслевого баланса, а также модели отдельных секторов: налогово-бюджетной и денежно-кредитной сферы и т.д. Развитием балансовых моделей являются модели с обратными связями.

Наиболее сложным вариантом математических моделей является совокупность моделей с участием человека. Эти модели создаются для принятия управленческих решений, моделированием процессов адаптации системы к изменениям внешней среды, рефлексивное поведение отдельных групп и участников. Методами этого уровня абстракции являются управление знаниями, экспертиза, экспертные оценки, обучение, накопление опыта и т.д.

В диаграмма экономико-математической модели приведена на рис.2



Рисунок 7. V диаграмма экономико-математической модели

## 5. Соотношение информационных и экономико-математических моделей

Свое воплощение в жизнь информационные и экономико-математические модели находят в информационных системах. Модели могут быть двух типов: абстрактная или практическая. Абстрактные модели помогают сформировать системное восприятие, а практические – симитировать реальность,

зафиксировать транзакции, обработать данные, и т.д. Чтобы абстрактная модель превратилась в информационную систему, она должна пройти через стадию реализации экономико-математического и информационного моделирования (рис.1 и 2).

В приложении информационная модель реализуется в основном в базе данных. Математические модели реализуются в виде программных компонент, которые выполняют обработку данных. Для соединения информационных моделей с программными компонентами могут быть использованы разные механизмы в зависимости от применяемого метода интеграции данных. Рассмотрим три категории интеграции: в контролируемой среде, в полу-контролируемой среде и в неконтролируемой среде.

В контролируемой среде, на уровне предприятий, данные как правило, связываются с объектным кодом посредством инструмента соединения сущностей с объектной моделью (Object Relational Mapping, ORM). Наиболее распространенными ORM являются Hibernate, ADO.NET Entity Framework, и т.д.

Для организации взаимодействия в полу-контролируемой среде, как правило, используются облачные сервисы, в которых технологиями доступа к данным выступают процедурные программные интерфейсы (API). Данные собираются в хранилище по технологии ETL, а информационная модель строится на основе данных локальных источников (Global-As-View, GAV).

В случае неконтролируемой среды технологиями организации данных выступают технологии стандартизированной формы описания ресурсов Resource Definition Framework, RDF. Доступ к данным осуществляется путем сериализации RDF формата, а для идентификации используется URI. Альтернативой хранения описанных данных произвольной структуры могут выступать XML- или граф-ориентированные базы данных. Доступ к данным в этом случае осуществляется с использованием специализированных языков запросов, таких как SPARQL, либо технологии работы с XML документами: XPath, XQuery. Информационная модель строится исходя из собственного представления предметной области, по принципу Local-as-View (LAV).

Информационные и экономико-математические модели информационной системы отражают разные срезы реального мира. При построении информационных систем в распределенной среде две категории моделей изменяют свою форму. В таблице 1. приведено соотношение информационных и экономико-математических моделей.

**Таблица 1 Соотношение экономико-математических и информационных моделей в экономических системах**

<b>Направление</b>	<b>Информационные модели</b>	<b>Экономико-математические модели</b>
<b>Характеристика</b>	Информационная	Интеллектуальность

<b>информационной системы</b>	полнота	системы
<b>Предназначение модели</b>	Поставка данных	Обработка данных
<b>Задача</b>	Отразить прошлое или настоящее состояние	Предсказать новые свойства или новые результаты явления
<b>Стадии управления</b>	Понимание	Прогнозирование, оптимизация, имитация
<b>Отражение объекта в модели (объекты, ситуации, явления, процессы)</b>	Сущность, атрибут, связь, тип данных	Система уравнений, арифметических отношений, геометрических фигур или их комбинация; моделирование процессов и оценка параметров
<b>Распределенные технологии</b>	Распределенные данные	Распределенные вычисления
<b>Технологии интеграции</b>	Интеграция данных: ETL, федерация, семантический веб	Имитационные методы: дискретно-событийные модели, агентские модели, системная динамика; Аналитические методы: балансовые модели, модели с обратной связью
<b>Стандартизация</b>	Стандартизация данных	Стандартизация облачных вычислений
<b>Сервисные (облачные) технологии</b>	Данные как сервис: поставка данных из реестров	Программные продукты как сервис: предоставление вычислительных методов как сервис
<b>Открытые технологии</b>	Открытые данные	Свободно распространяемые библиотеки ПО

Основой современных экономических систем являются качественные данные. Если система содержит качественные данные, то применением таких простых математических методов как группировка и расчет элементарных статистик подготавливается значительная часть экономических отчетов, выполняется визуализация данных. Это позволяет пройти этап понимания ситуации, следующие этапы анализа данных, связанные с симуляцией, прогнозированием, оптимизацией, являются более математически насыщенными.

Информационные и экономико-математические модели в некоторой степени являются независимыми друг от друга, то есть ортогональны при отражении их в пространстве. Каждая из этих моделей может иметь разную степень сложности. При совместной реализации двух категорий моделей на практике наилучшем вариантом развития является возрастание степени сложности экономико-математической модели при росте информационной наполненности модели. В реальной ситуации мы имеем другую зависимость (Рис.3): при возрастании сложности математической модели происходит снижение информационного наполнения, и наоборот – при росте информационного наполнения снижается уровень математической модели, которую мы можем реализовать на имеющихся данных.



Рисунок 8. Соотношение экономико-математических и информационных моделей

С ростом количества поставщиков данных появляется возможность реализации экономико-математических моделей не на выборке, а на всей совокупности данных. Развитие информационных технологий на сегодняшнем этапе таково, что позволяет реализовать выход на макроэкономическую модель путем сбора информации с микроуровня, возможно элементами имитационной модели. Это позволяет получить прозрачную сквозную картину от микроуровня к мезоуровню и далее на макроэкономику. При реализации таких проектов (ГАС «Управление», в области здравоохранения) значения показателей, получаемых агрегацией, часто существенно отличаются от тех, которые получаются по текущей методике с применением методов оценок.

Возможность агрегирования данных микроэкономики до макроуровня актуальна для многих секторов экономики России. Например, в финансовой системе России отсутствует целый ряд элементов финансового счета, не решена задача составления баланса финансовых инструментов и институциональных балансов, что не позволяет принимать качественные решения. Потребность в таких инструментах испытывает не только Банк России, но и силовые структуры.

## **6. Выводы и направления дальнейших исследований**

В статье представлено соотношение двух категорий моделей, описывающих экономическую деятельность: информационных и экономико-математических моделей. Каждая из категорий включает широкий класс типов моделей и методов их реализации. Сделано предположение, что на данном этапе развития экономических приложений информационные модели занимают первостепенное значение. Наличие качественных данных позволит пройти этап «Понимание» в управленческой практике.

Современные информационные системы работают в распределенной среде и предполагают интеграцию данных и вычислительных методов. Комплексная интеграция данных и вычислений реализуется на уровне локального предприятия для взаимодействия отдельных функций. В мировой практике существуют примеры реализации информационной модели национального масштаба на уровне обмена документами, прорабатывается модель обмена с использованием семантических методов.

Использование индустриальных методов поставки данных способно перевести управление на более качественный уровень. Одним из важных элементов развития информационной модели является методология создания логической модели данных для разных секторов экономики. Вариантом такой методологии могут быть шаблоны для отражения в модели данных основных направлений, таких как базовые объекты предметной области, их транзакции, роли участников, пространственно-временные характеристики, а также показатели мотивации экономической деятельности. Поставщиками данных отдельных элементов такой модели должна быть федеральная информационная инфраструктура.

## Литература

**Грачева М.В., Черемных Ю.Н. и Туманова Е.А.** Моделирование экономических процессов [Книга]. - Москва : ЮНИТИ-ДАНА, 2013.

**Колин К.К.** Философия информации и современное научное мировоззрение [Конференция] // Современные информационные технологии и ИТ-образование. - Москва : [б.н.], 2014.

**Месарович М. и Такахара Я.** Общая теория систем: математические основы [Книга]. - Москва : Мир, 1978.

**Минкомсвязь РФ** Технологический портал СМЭВ [В Интернете]. - 12 1 2015 г.. - <http://smev3.gosuslugi.ru/portal/>.

**Самарский А.А. и Михайлов А.П.** Математическое моделирование: Идеи. Методы. Примеры [Книга]. - Москва : Физматлит, 2001.

**Dietz Jan L.G.** Enterprise Ontology Theory and Methodology [Отчет]. - [б.м.] : Springer, 2006.

**ISA** European Interoperability Reference Architecture (EIRA) [В Интернете]. - 10 4 2015 г.. - <https://joinup.ec.europa.eu/asset/eia/description>.

**ОИП** National Information Exchange Model [В Интернете]. - 22 2 2015 г.. - <https://www.niem.gov/>.

**OMG** Semantic Information Modeling for Federation (SIMF RFP) [В Интернете]. - 2011 г.. - 12 8 2014 г.. - <http://tinyurl.com/SIMFrfp>.

**OMG UPDM** Unified Profile for DoDAF, MODAF, and NAF [В Интернете]. - 2 2 2015 г.. - <http://www.omg.org/spec/UPDM/>.

**Statistical Working Group** Statistical Data and Metadata Exchange [В Интернете]. - 23 3 2015 г.. - <http://sdmx.org/>.

# РАЗВИТИЕ ПРОЦЕССОВ АВТОМАТИЗАЦИИ ПРИ ФОРМИРОВАНИИ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ПОДДЕРЖКИ МЕЖДИСЦИПЛИНАРНЫХ НАУЧНЫХ ИССЛЕДОВАНИЙ

И.А. Мбого<sup>1,2</sup>, Д.Е. Прокудин<sup>2</sup>, А.В. Чугунов<sup>1</sup>

<sup>1</sup> Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО)

<sup>2</sup> Санкт-Петербургский государственный университет

**Аннотация.** В настоящее время применение большого числа разнородных информационных систем и программных платформ для поддержки научной деятельности выявляет проблемы, связанные с отсутствием унификации обмена метаданными между ними, а также ограниченность или полное отсутствие автоматизации при обмене между ними информационными потоками. Статья посвящена рассмотрению решению задач автоматизации информационных потоков при реализации комплексного подхода формирования информационного пространства научных исследований на примере междисциплинарного направления изучения технологий информационного общества.

**Ключевые слова:** информационное пространство научных исследований; информатизация научной деятельности; информационно-коммуникационные технологии; автоматизация; метаданные.

**Введение.** Тенденции развития науки в концепции постнеклассической науки базируются на массовых процессах информатизации, внедрения в практику научных исследований информационно-коммуникационных технологий. Это, в свою очередь, порождает различные инициативы: «открытая наука» [4], «открытый доступ» [1, 3, 7] к результатам научных исследований и т.д. В соответствии с этими тенденциями происходит виртуализация различных видов научной деятельности, переход их в гибридную (аналого-цифровую) и даже полностью электронную форму существования [2, 8]: информационно-поисковая деятельность, представление и распространение результатов научно-исследовательской деятельности, научная коммуникация и т.д. На сегодняшний день для этих целей в научной среде используется огромное количество различных технологических решений. Одной из важных проблем является то, что среди существующих систем нет таких, которые могли бы комплексно охватить все виды научной деятельности, а разработка такой комплексной системы поддержки научных исследований требует достаточно больших затрат (организационных, инвестиционных и т.д.) и тщательной концептуальной и



структурной разработки [5]. Поэтому объективное наличие в эксплуатации огромного числа разнородных информационных систем ставит такие проблемы как: отсутствие единой системы аутентификации (и, следовательно, идентификации личности учёного); отсутствие или ограниченность реализации механизмов автоматизации при взаимодействии систем на уровне передачи информационных потоков (например, подача рукописи статьи или передача метаданных); отсутствие единых стандартов обмена метаданными (или даже отсутствие такой возможности) и т.п.

**Средства стандартизации и автоматизации в информационных системах для научных исследований.** Несмотря на существующие проблемы разнородности в современных технологических решениях разработаны стандарты, которые учитывают мировые инициативы «свободного доступа». К основным из них можно отнести протокол обмена метаданными OAI-PMH и стандарт Дублинского ядра для простого открытия ресурса (Dublin Core), отвечающий за совместимость с электронными репозиториями, архивами и иными источниками информации. Их поддержка реализована в таких широко распространённых программных платформах как:

DSpace (<http://www.dspace.org>), которая разрабатывалась для организации электронных библиотек, поэтому в основном это решение используется для создания различных институциональных репозиториях разнородных документов.

ePrints (<http://www.eprints.org/uk/>), вышедшая из движения Scholarly Communication («Общение в мире науки») и поэтому используется для построения репозиториях научных статей, препринтов и отчётов результатов научных исследований. При этом ПО ePrints оптимизировано для обеспечения доступа к самостоятельному депонированию автором своих материалов.

Open Journal System (OJS) – представляет собой онлайн-издательскую систему полного цикла, которая реализует концепцию «электронного издательства».

В этих системах данные стандарты реализованы на уровне провайдера, что позволяет решать задачу актуализации и оперативного распространения результатов научных исследований через размещение метаданных в основных агрегаторах научной информации. К ним, например, относятся следующие информационные системы:

The OAIster® database (<http://oaister.worldcat.org>) – один из самых мощных мировых агрегаторов, который содержит более 30 млн. записей мета данных публикаций более чем 1,5 тыс. организаций-участников, предоставляющих свои публикации по принципу «открытого доступа». Эта разработка Мичиганского университета в настоящее время поддерживается и развивается Online Computer Library Center, Inc. (OCLC), являющейся научно-исследовательской организацией, общественной целью которой является расширение доступа к мировой информации и сокращение расходов на информацию;

Registry of Open Access Repositories (ROAR - <http://roar.eprints.org>, созданный в Саутхемптонском университете. В настоящее время в ROAR зарегистрированы 3,3 тыс. архивов, в том числе всего 53 из России;

Directory of Open Access Repositories (OpenDOAR - <http://www.opendoar.org>) создан и поддерживается Ноттингемским университетом (Великобритания). В настоящее время OpenDOAR содержит 2845 архива, из которых 22 являются российскими;

Directory of Open Access Journals (DOAJ - <http://doaj.org>) – каталог научных журналов открытого доступа, идея создания которого получила одобрение в 2002 г. на Первой Скандинавской конференции по проблемам научных коммуникаций в г. Лунде (Швеция).

**Подход к развитию процессов автоматизации в информационном пространстве научных исследований.** В рамках развития комплексного информационного пространства междисциплинарного направления научных исследований технологий информационного общества в качестве базового элемента была выбрана ежегодная научная конференция «Интернет и современное общество» (<http://ims.ifmo.ru/>). За 14 лет проведения конференции сформирована весьма объемная информационная и методическая база, в которой заинтересовано сообщество исследователей из большинства российских регионов.

Для аккумуляции материалов конференции построена информационная система на базе OJS (<http://ojs.ifmo.ru/index.php/IMS/>), в которой уже размещены материалы конференций с 2011 по 2014 годы, представляющие собой научные и научно-методические публикации по вопросам развития информационного общества, создания информационных ресурсов и систем для поддержки научных исследований в сфере гуманитарных наук [6]. В данном случае выбор OJS основан на том, что эта платформа позволяет решить комплекс задач: 1) построения репозитория; 2) автоматизация приёма материалов докладов (за счёт реализованного стандартного процесса); 3) предоставление метаданных публикаций во внешние агрегаторы (поддержка протокола OAI-PMH на уровне провайдера и формата Dublin Core); 4) автоматизация сохранения цифрового контента (модуль поддержки архивирования в Private LOCKSS Networks, созданную (начиная с версии 2.4.5) в рамках развития Public Knowledge Project (<http://pkp.sfu.ca>); 5) подготовка метаданных для предоставления во внешние информационные системы (агрегаторы), не поддерживающие средства автоматизации (модули экспорта метаданных в Erudit, определенном в виде DTD; CrossRef XML; PubMed XML для индексирования MEDLINE; XML для архивации в DOAJ).

Была проведена регистрация системы в агрегаторах OAIster ([http://www.worldcat.org/search?q=on:DGCNT+http://ojs.ifmo.ru/index.php/index/oai+IMS+RUITM&qt=results\\_page](http://www.worldcat.org/search?q=on:DGCNT+http://ojs.ifmo.ru/index.php/index/oai+IMS+RUITM&qt=results_page)), Соционет (<https://socionet.ru/collection.xml?h=repec:rus:ims000>) по протоколу OAI-PMH и

Академии Гугл (с регистрацией по типу системы OJS), а также настроена автоматическая синхронизация метаданных с ними. Кроме этого в системе Артикулус были подготовлены метаданные сборников статей за 2011, 2012, 2013 и 2014 года и размещены в Научной электронной библиотеке (<http://elibrary.ru>).

Информационное пространство дополняют сайт конференции «Интернет и современное общество» (<http://ims.ifmo.ru>) и тематическая электронная библиотека «Электронное государство» (<http://library.egov.ifmo.ru>). В библиотеку включаются избранные материалы, в том числе, из материалов конференций. Библиотека выполнена на базе системы управления контентом Drupal и также использует механизм обмена метаинформацией OAI-PMH.

**Заключение.** При построении комплексного информационного пространства разработчики столкнулись с проблемами многочисленного ручного ввода метаданных подаваемых на конференцию материалов – как при первичной подаче, так и при вводе данных в Научную электронную библиотеку. Поэтому задача, которая будет решаться на следующем этапе развития информационного пространства заключается, с одной стороны, в разработке шаблона для авторов, который позволит переводить метаданные в XML-форматы основных информационных систем, составляющих информационное пространство, а, с другой стороны, – в разработке модуля для OJS автоматического импорта метаданных сборников научных статей для размещения в Научной электронной библиотеке (РИНЦ).

## ЛИТЕРАТУРА

1. Будапештская инициатива «Открытый Доступ» // Budapest Open Access Initiative. Russian Translation. URL: <http://www.budapestopenaccessinitiative.org/translations/russian-translation> (дата обращения: 24.05.2015).
2. Елизаров А.М., Зуев Д.С., Липачёв Е.К.. Сервисы поддержки жизненного цикла электронных научных публикаций // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров: Труды Международной суперкомпьютерной конференции (22-27 сентября 2014 г., г. Новороссийск). – М.: Изд-во МГУ, 2014. С. 436-438.
3. Земсков А.И., Шрайберг Я.Л. Системы открытого доступа к информации: причины и история возникновения // Науч. и техн. б-ки.- 2008. №4. URL: [http://ellib.gpntb.ru/subscribe/ntb/2008/4/ntb\\_4\\_2\\_2008.htm](http://ellib.gpntb.ru/subscribe/ntb/2008/4/ntb_4_2_2008.htm) (дата обращения: 24.05.2015).
4. Паринов С.И. Развитие электронных библиотек – путь к Открытой Науке // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской научной конференции RCDL'2009. Петрозаводск: КарНЦ РАН, 2009. С. 225-234. URL: [http://rcdl.ru/doc/2009/225\\_234\\_Invited-2.pdf](http://rcdl.ru/doc/2009/225_234_Invited-2.pdf) (дата обращения: 24.05.2015).

5. Прокудин Д.Е. Проектирование и реализация комплексной информационной системы поддержки научных исследований // Технологии информационного общества в науке, образовании и культуре: сборник научных статей. Труды XVII Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2014), Санкт-Петербург, 19 – 20 ноября 2014 г. С. 31-36.
6. Мбого И.А., Чугунов В.А., Прокудин Д.Е. Формирование информационного пространства междисциплинарного научного направления: подходы и решения // Межотраслевая информационная служба. 2015. №1 (170). С. 36-44. URL: <http://elibrary.ru/item.asp?id=23082608> (дата обращения: 24.05.2015).
7. Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Feb 28th - Mar 1st, 2005, University of Southampton, UK. URL: <http://www.eprints.org/events/berlin3/outcomes.html> (дата обращения: 24.05.2015).
8. Elizarov A.M., Zuev D.S., Lipachev E.K. Electronic scientific-journal management systems // Scientific and Technical Information Processing. 2014. V. 41, No 1. P. 66-72.

# ОРГАНИЗАЦИЯ ВНЕШНЕГО ПОЧТОВОГО SMTP-СЕРВЕРА В НАУЧНОЙ ОРГАНИЗАЦИИ

Г.М. Михайлов, Ю.П. Рогов, А.М. Чернецов

*Вычислительный центр им. А.А. Дородницына Российской академии наук*

В данной работе рассматривается вопрос создания почтового smtp-сервера научной организации, доступного для обращения вне территории организации. Это часто требуется для отправки электронной почты сотрудниками вне рабочего места, например, находясь в командировке или используя домашний компьютер.

Есть различные варианты решения поставленной задачи:

- 1) открыть доступ к smtp-серверу для «внешнего мира»;
- 2) использовать web-интерфейс, открытый для «внешнего мира»;
- 3) создать дополнительный smtp-сервер, доступный извне.

В первом случае либо возникает ситуация “open relay”, когда smtp-сервер может использоваться кем угодно, в том числе и для организации спам-атак [1], либо используется авторизация и (или) шифрование при отправке электронной почты. При этом необходимо обеспечить доступ пользователю во внутреннюю сеть организации. Если политика безопасности организации не позволяет таких действий, то smtp-сервер необходимо выводить в отдельный изолированный сегмент, не имеющий доступа в основную сеть организации.

Использование web-интерфейса снимает проблему доступа, однако имеет следующие недостатки:

- если почтовая служба параллельно использует протокол **pop3**, то невозможно хранить отправленную почту;
- если прерывается связь с сервером, то нет и доступа к локальной копии почты на рабочем месте пользователя.

По этим причинам пользователи предпочитают работу с почтовым клиентом, не пользуясь **web-интерфейсом**.

В ВЦ РАН долгое время была реализована следующая схема: **smtp-сервер** во внутренней сети, недоступный для отправки сообщений сотрудниками вне рабочих мест. Впоследствии к серверу был «подключен» web-интерфейс, однако он доступен также исключительно из сети организации. Поэтому вопрос доступа к почтовым ресурсам оставался актуальным.

Электронная почта ВЦ РАН организована на базе **sendmail 8.13.8/Solaris10**.

В качестве **smtp-сервера** используется **sendmail 8.11.6/Solaris 8 SPARC**. К большому сожалению, после покупки компании Sun Microsystems компанией Oracle, произошли значительные изменения в лицензировании операционной системы Solaris. В результате обновления операционной системы (ОС)

доступны только по подписке, а сама операционная система стала платной. Поэтому было принято решение о смене платформы для *smtp-сервера*.

Для реализации нового *smtp-сервера* была выбрана платформа **CentOS 6.6** с запуском на действующей в ВЦ РАН системе виртуализации **Microsoft HyperV** [2]. Использование виртуализации обусловлено двумя причинами:

- 1) используется меньше физического оборудования и меньше затрат на его обслуживание;
- 2) в случае проблем (взлома) проще погасить виртуальную машину, чем отключать физическое оборудование.

В качестве *smtp-сервера* был выбран **Sendmail версии 8.14.4**.

Стандартная конфигурация **sendmail** предполагает, что сервер может использоваться как для отправки почтовых сообщений, так и для их приема. В нашем случае необходимо исключить возможность приема сообщений в принципе. Для решения задачи производится перенастройка MTA-агента:

вместо

```
define(`confAUTH_OPTIONS', `A p')dnl
```

необходимо использовать

```
define(`confAUTH_OPTIONS', `A y')dnl
```

Как всегда, для любого «нормального» сервера необходимо позаботиться о настройках, блокирующих спам и снижающих нагрузку. Для этого применяется несколько механизмов:

```
FEATURE(`relay_based_on_MX')dnl
```

```
define(`confMAX_DAEMON_CHILDREN', `10')dnl
```

```
define(`confCONNECTION_RATE_THROTTLE', `5')dnl
```

использование `access_db`

```
FEATURE(`blacklist_recipients')dnl
```

```
FEATURE(`dnsbl', `xbl.spamhaus.org', `You are trying to send spam - see ?"$&{client_addr}')dnl
```

```
FEATURE(`dnsbl', `sbl.spamhaus.org', `You are trying to send spam - see ?"$&{client_addr}')dnl
```

```
FEATURE(`dnsbl', `cbl.abuseat.org', `Rejected - see ?"$&{client_addr}')dnl
```

```
FEATURE(`dnsbl', `relays.ordb.org', `Rejected - see ?"$&{client_addr}')dnl
```

```
define(`confTO_CONNECT', `30s')dnl
```

– установить задержку на ответ сервера клиентам.

Для обеспечения защищенного соединения необходимо иметь сертификат X.509. Этот сертификат может быть как самоподписанным (что часто бывает даже в весьма крупных компаниях), либо выпущен каким-либо центром сертификации. В ВЦ РАН используется сертификат, выпущенный корпоративным центром сертификации.

Работа представлена в рамках выполнения темы: **«Локальная информационно-вычислительная система ВЦ РАН – среда для фундаментальных и прикладных исследований учреждения»**

## ЛИТЕРАТУРА

1. Копытов М.А., Рогов Ю.П. Электронная почта. Администрирование и проблемы надежности. Тезисы доклада в сборнике Всероссийской научной конференции "Научный сервис в сети Интернет" (г.Новороссийск, 23-28 сентября 2002 года). – М.: Изд-во МГУ, 2002. – С. 128-129.
2. Михайлов Г.М., Рогов Ю.П., Чернецов А.М. Система хранения данных в локальной компьютерной сети ВЦ РАН. Тезисы доклада в трудах Международной суперкомпьютерной конференции "Научный сервис в сети Интернет: экзафлопсное будущее" (г.Новороссийск, 19-24 сентября 2011 г.). – М.: Изд-во МГУ, 2011. – С. 488-489.  
URL: <http://agora.guru.ru/abrau2011/pdf/488/pdf>

# УЧЕТ ДАННЫХ НАБЛЮДЕНИЙ ДЛЯ АНАЛИЗА ДИНАМИЧЕСКИХ ПРОЦЕССОВ В ОКЕАНСКОЙ МОДЕЛИ ЦИРКУЛЯЦИИ

Г.М.Михайлов<sup>1</sup>, К.П.Беляев<sup>2,1</sup>, А.Н.Сальников<sup>3,1</sup>, Н.П.Тучкова<sup>1</sup>, И.Кирхнер<sup>4</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки  
Вычислительный центр им. А. А. Дородницына РАН

<sup>2</sup> Федеральное государственное бюджетное учреждение науки  
Институт океанологии им. П.П. Ширшова РАН

<sup>3</sup> Федеральное государственное образовательное учреждение высшего  
профессионального образования "Московский государственный университет  
имени М.В.Ломоносова", факультет ВМиК

<sup>4</sup> Свободный университет Берлина, факультет Метеорологии

В работе рассматривается вариант адаптации модели MPIESM(ESMAM6/MPIOM/JSBACH/HAMOCC). Модель MPIESM [1, 2] представляет собой расширенную версию совместной климатической модели COSMOS (ESMAM5/MPIOM), разработанную в европейском научном профессиональном сообществе климатических исследований и реализованную в Немецком вычислительном центре климата DKRZ(<http://www.dkrz.de/dkrz-en>). Весь комплекс MPIESM включает модель океана с учетом подмодели динамики льда (MPI-OM), модель атмосферы (ESMAM6), модель земной биосферы (JSBACH), блок биохимии океана (HAMOCC) и блок взаимодействия моделей (OASIS3). MPIESM является совместной моделью, где все процессы осуществляются в параллельном режиме при взаимодействии океанской, атмосферной частей и модели поверхности земли посредством обмена энергией, импульсов и других параметров. Процессы передачи энергии, импульса, параметров воды и углекислого газа реализуются с помощью блока взаимодействия моделей (OASIS3). Тестирование модели проводилось с учетом рекомендаций из отчета МГЭИК (Межправительственная группа экспертов по изменению климата) и учитывает моделирование углеродного цикла. Результаты тестирования и исходные коды предоставляются научному сообществу на лицензионной основе в исследовательских целях.

Для адаптации модели на вычислительном комплексе «Ломоносов» был проведен анализ функциональности модели и параллельного взаимодействия блоков модели. При компиляции были изменены ключи и настройки библиотек, скрипты для запуска. Также были подготовлены исходные данные для численных экспериментов piControl\_r1i1p1-LR, historical\_r1i1p1-LR, amip\_r1i1p1-LR, sstClim\_r1i1p1-LR. Эти данные для сравнения и верификации новых реализаций совместных моделей включают:

– историческое моделирование климата (с 1850 г. до 2005 г.);



- прогнозирование климата с различными сценариями (от 2100 г. до 2300 г.);
- прогноз климата на каждые десять лет;
- учет роли углеродного цикла для изменений климата;
- моделирование более отдаленного прошлого, например последнего ледникового максимума или от 850 г. до 1850 г.

Для запуска модели были адаптированы коды для параллельной платформы суперкомпьютера «Ломоносов»; использовалась гибридная параллельная структура на основе MPI1 и OpenMP библиотек; реализовывались новые версии скриптов для запуска всех блоков модели одновременно в параллельном режиме; для верификации расчетов на новой платформе последовательно реализуются этапы тестовых экспериментов. При отладке модели на новой кластерной платформе для верификации было проведено сравнение результатов вычислительных экспериментов с данными по изучению изменения климата, предоставляемыми разработчиками модели MPIESM.

На рис. 1 показаны основные блоки модели MPIESM: блок ECHAM6 – модель динамики атмосферы, блок MPIOM – модель динамики океана, включающая динамику льда, блок OASIS – специальный интерфейс обмена океан-земля-атмосфера и блок HAMOCC – модель газообмена, включающая динамику карбонных соединений, а также блок JSBACH – модель стоков крупнейших рек планеты.

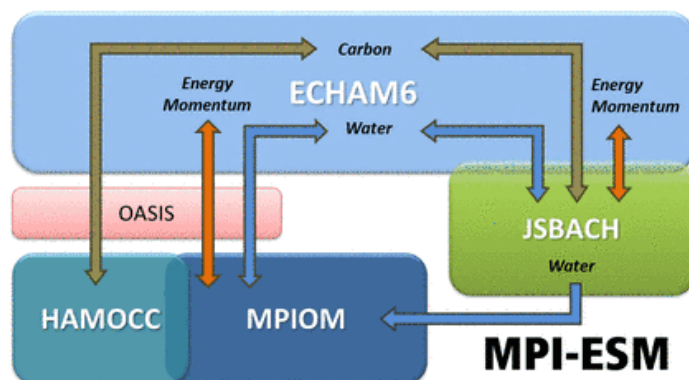


Рис. 1. Структура и взаимосвязи модели MPIESM

В рамках проекта также разрабатывался новый метод усвоения данных наблюдений, с применением теории диффузионных процессов и стохастических дифференциальных уравнений [3, 4, 5]. Метод основан на минимизации функционала следующего вида

$$L(K, \phi) = KQK' + [(I + KH)\Lambda]\phi, \quad (1)$$

где:

$K$  – неизвестная и подлежащая определению матрица размерности  $r \times n$ ,

$Q$  – ковариационная матрица наблюдений размерности  $n \times n$ , предполагается известной,

$H$  – известная матрица размерности  $n \times r$ , имеющая смысл проекционной матрицы из пространства модели размерности  $r$  в пространство наблюдений размерности  $n$ ,

$I$  – единичная матрица размерности  $r \times r$ ,

$\Lambda$  – оператор модели, имеющий размерность  $r \times 1$ ,

$\varphi$  – вектор множителей Лагранжа, имеющий размерность  $1 \times r$ .

Штрих вверху вектора или матрицы обозначает ее транспонирование.

Минимизация функционала (1) имеет физический смысл определения минимума дисперсии или диффузии, задаваемой матрицей  $KQK'$  при условии известного или определяемого из наблюдений вектора сноса  $(I + KH)\Lambda$ .

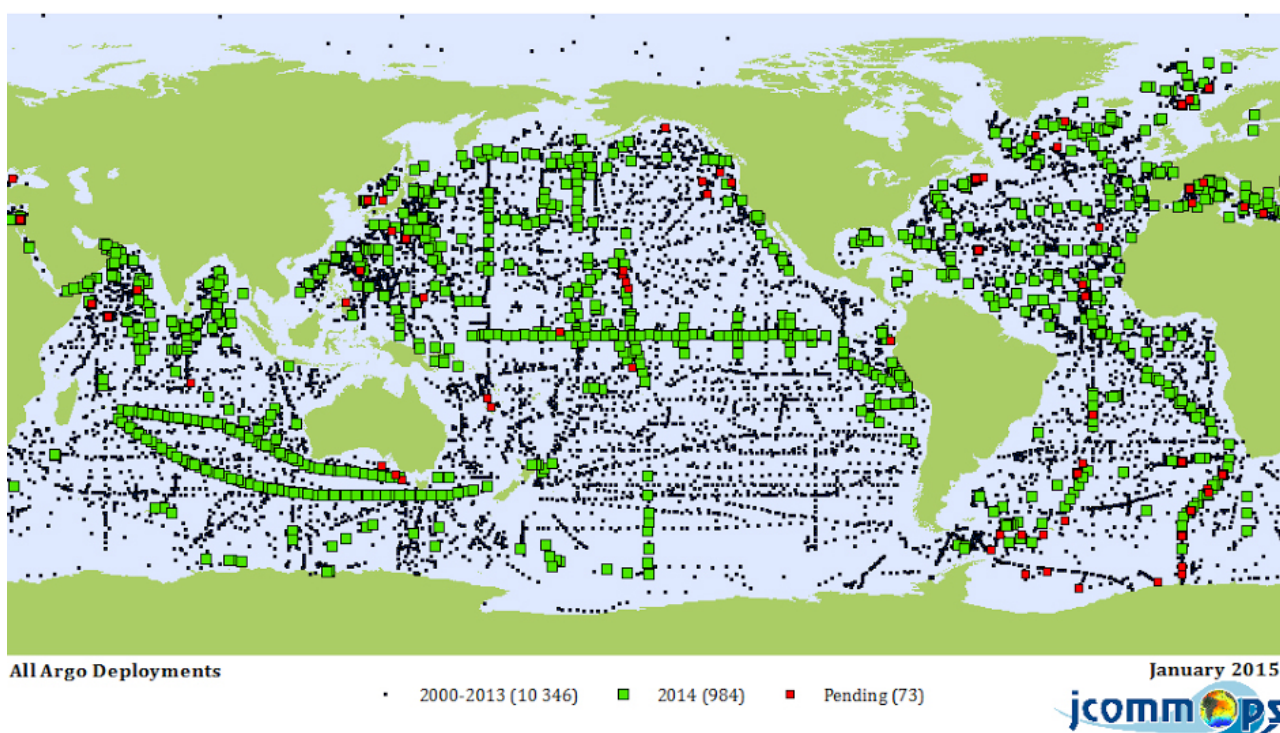


Рис. 2. Карта дрейфтеров АРГО в мировом океане

В проекте предусмотрена состыковка модели MPIESM и метода усвоения данных, эксперименты с усвоением данных наблюдений со спутников (уровня океана и поверхностной температуры воды), а также данных дрейфтеров АРГО (<http://www.argo.net/>), рис. 2, и анализ результатов данных экспериментов. Кроме того, предусматривается сравнение авторских методов и моделей с исследованиями других авторов.

Работа поддерживается РФФИ (проекты 14-07-00037, 14-05-00363).

## ЛИТЕРАТУРА

1. Haak H. Simulation of Low-Frequency Climate Variability in the North Atlantic Ocean and the Arctic, Volume 1. Max Planck Institute for Meteorology, 2004.
2. Wetzel P., Haak H., Jungclaus J., Maier-Reimer E. The Max-Planck-Institute Global Ocean/Sea-Ice. Model URL: [http://www.mpimet.mpg.de/fileadmin/models/MPIOM/DRAFT\\_MPIOM\\_TECHNICAL\\_REPORT.pdf](http://www.mpimet.mpg.de/fileadmin/models/MPIOM/DRAFT_MPIOM_TECHNICAL_REPORT.pdf) (Model MPI-OM. Technical report).
3. Беляев К.П., Тучкова Н.П., Кирхнер И. Метод коррекции модельных расчетов по данным измерений, основанный на диффузионном приближении, и его применения для анализа гидрофизических характеристик // Математическое моделирование, 2009, т. 21, № 3, с. 53-68.
4. Беляев К.П., Тучкова Н.П., Кубаш У. Реакция совместной модели "океан-лед-атмосфера" при усвоении данных наблюдений в тропической зоне Тихого океана // Океанология, 2010, т. 50, № 3. С. 334-344.
5. Беляев К.П., Тучкова Н.П. Предельные распределения для характеристик при усвоении данных наблюдений в стационарном режиме // Информатика и её применения, 2015, т. 3. Вып. 2. С. 19-24.

# МОДЕЛИ УПРАВЛЕНИЯ КОЛЕБАТЕЛЬНЫМИ ПРОЦЕССАМИ И ИХ ТЕРМИНОЛОГИЧЕСКОЕ ОПИСАНИЕ В ИНФОРМАЦИОННОЙ СРЕДЕ

Е.И. Моисеев<sup>1,2</sup>, А.А. Муромский<sup>1</sup>, Н.П.Тучкова<sup>1</sup>, К.Б. Теймуразов<sup>1</sup>

*1 Федеральное государственное бюджетное учреждение науки  
Вычислительный центр им. А. А. Дородницына РАН*

*2 Федеральное государственного образовательное учреждение высшего  
профессионального образования "Московский государственный университет  
имени М.В.Ломоносова", факультет ВМиК*

## ВВЕДЕНИЕ

В работе дается вариант представления математических предметных областей в интернет. В качестве информационной модели для единицы записи выбрана статья тезауруса. Тезаурус позволяет по указанному в запросе одному термину и отношениям выявить и другие термины, связанные с данным термином этими отношениями, а значит получить дополнительные вторичные документы [1-3]. Реализация такой возможности особенно важна для поиска в режиме избирательного распространения информации. (ИРИ, ГОСТ 7.73-96 ). Реализация схемы показана на примере уравнений с частными производными (УЧП). Выбранная предметная область содержит обширный класс задач и актуальных приложений. Одно из основных уравнений математической физики - волновое уравнение (Даламбера уравнение, уравнение бегущих волн, уравнение колебаний струны). Работы адресата по данной тематике используются для составления списка ключевых слов предметной области адресата. Используются классические источники, представленные в списке литературы, в том числе списки русскоязычной литературы, которые, не всегда цитируются в иностранных источниках, и авторские разработки [4-8].

Волновое уравнение является одним из важнейших уравнений с частными производными второго рода, которые применяют для описания волн различной природы. Эти физические процессы возникает в таких областях как акустика, электромагнетизм и гидрогазодинамика. Исторически проблемы вибрации струн, таких как струны музыкальных инструментов были изучены Даламбером (Jean le Rond d'Alambert), Эйлером (Leonard Euler), Бернулли (Daniel Bernoulli), Лагранжем (Joseph-Louis Lagrange). В 1746 г. Даламбер открыл одномерное волновое уравнение, а через 10 лет Эйлер открыл трехмерное волновое уравнение. Основное волновое уравнение - линейное дифференциальное уравнение.

## О ПРЕДСТАВЛЕНИИ ПОНЯТИЙ ПРЕДМЕТНОЙ ОБЛАСТИ

Предлагается один из вариантов идентификации физических процессов в информационно-поисковой среде [9, 10]. Физические и технические процессы

исследуются с помощью моделей математической физики (уравнения, задачи, методы, решения), идентифицируются терминами без математической символики и сопровождающими эти термины буквенно-цифровыми кодами – идентификаторами<sup>5</sup>.

Структура термина:

<слово: задача><название физического/технического процесса с указанием основных характеристик, свойств процесса и материального объекта без математической символики>

Структура идентификатора для физических/технических задач:

<Pr><PDE><XXXX>, PrPDEXXXX.

Буква <Pr> указывает на то, что данный термин является названием задачи (Problem). Аббревиатура <PDE> (Partial Differential Equations) показывает, что задачу решают средствами математической физики (МФ).

<XXXX> - уникальный десятичный код ( $0001 \leq XXXX \leq 9999$ ).

Приведем список некоторых задач, в определении которых использована терминология из источников [4-8]. Заметим, что в литературе один и тот же физический процесс нередко называется по-разному. Это замечание касается терминологии в рамках предметной области. Указанные ниже термины уникальны: по сути дела это - термины-дескрипторы в данной предметной области.

Список физических задач. Идентификация физических процессов.

1. Задача о свободных поперечных колебаниях однородной бесконечной струны |PrPDE0001|
2. Задача о свободных продольных колебаниях однородного стержня |PrPDE0002|
3. Задача о свободных электрических колебаниях в проводах |PrPDE0010|
4. Задача о колебаниях газа в бесконечной цилиндрической трубке |PrPDE0031|
5. Задача о стационарном тепловом состоянии однородного тела |PrPDE0040|
6. Задача о распространении тепла в ограниченном стержне, на концах которого происходит свободный теплообмен с окружающей средой |PrPDE0052|
7. Задача оптимального граничного управления смещением на одном конце струны при свободном втором ее конце за произвольный достаточно большой промежуток времени |PrPDE0100|

---

<sup>5</sup> Идентификация здесь и далее – установление соответствия распознаваемого предмета, объекта, явления своему образу в виде знака, обычно именуемого идентификатором.

8. Задача о деформировании изотропного линейного упругого стержня, находящегося под действием собственного веса [PrPDE0150]

9. Задача о растяжении одноосного составного упругого тела с учетом масштабных эффектов в окрестности границы контактов [PrPDE0160]

Можно отметить также тенденции, характерные для терминологических обозначений различных видов уравнений с частными производными:

а) в наименовании PDE указывается 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

Систематизируем сведения, которыми можно располагать при формировании статьи тезауруса в виде таблицы 1.

Таблица 1. Информация о математической задаче предметной области МФ.

Основные сведения
Терминологическое обозначение – название данной задачи дескриптор данной задачи: [<Название задачи><Дескриптор PDE> <Информация о физическом процессе> <Сведения о материальной объекте>]
Термины- синонимы
Литература
<u>Математическая запись задачи</u>
<ul style="list-style-type: none"><li>○ Дескриптор PDE. Идентификатор PDE с указанием типа PDE</li><li>○ Запись PDE в декартовой системе координат</li><li>○ Запись PDE в используемой системе координат (отличной от декартовой)</li><li>○ Начальные условия</li><li>○ Граничные условия. Однородные, неоднородные условия</li><li>○ Интерпретация независимых переменных</li></ul>
<u>Корректность постановки задачи</u>
<ul style="list-style-type: none"><li>○ Утверждение о существовании и единственности решения задачи</li><li>○ Непрерывная зависимость решения задачи от исходных данных (условий)</li></ul>
<u>Решение задачи</u>
<ul style="list-style-type: none"><li>○ Точное решение задачи в виде формульной зависимости (аналитическое решение). Название точного решения</li><li>○ Приближенное решение задачи (численное решение). Название приближенного решения.</li><li>○ Табличное решение задачи.</li></ul>
<u>Методы решения задачи</u>
<ul style="list-style-type: none"><li>○ Методы точного решения</li><li>○ Методы приближенного решения</li><li>○ Методы табличного решения</li></ul>
<u>Связи</u>
<физические   технические задачи> → <задача предметной области МФ> :

Схематически представить родовидовые связи для УЧП второго порядка можно, например, как это сделано на рис.1

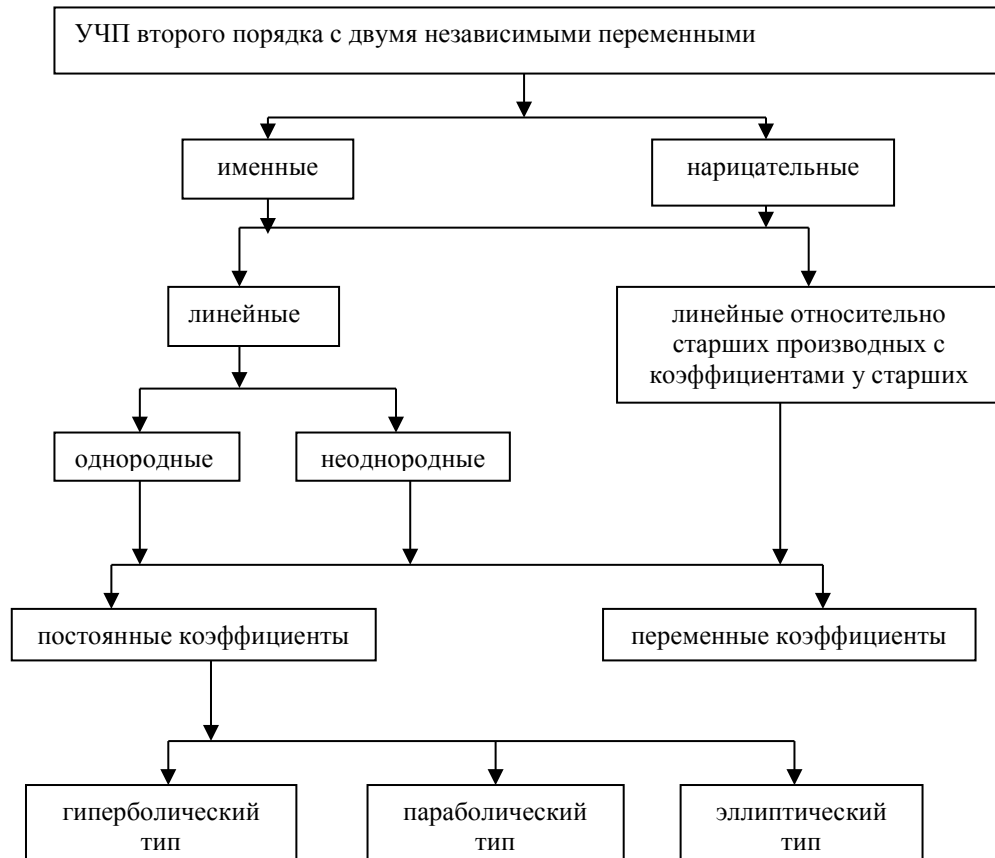


Рис.1. Схема графического представления линейного/линейного относительно старших производных УЧП(PDE) второго порядка с двумя независимыми переменными

## ВОПРОСЫ РЕАЛИЗАЦИИ

Представление тезауруса для волнового уравнения может быть следующим:

- Уравнения математической физики (Mathematical Physics Equations)
  - Волновое уравнение (wave equation)
  - синоним: Даламбера уравнение (D'Alembert equation)

синоним: уравнение бегущих волн (running-waves equation)

синоним: уравнение колебаний струны (string oscillation equation )

●●● волновое одномерное уравнение (one-dimensional wave equation)

- волновое одномерное уравнение (one-dimensional wave equation)
- волновое одномерное однородное уравнение свободных поперечных колебаний бесконечной однородной струны
- волновое одномерное однородное уравнение свободных поперечных колебаний ограниченной однородной струны
- волновое одномерное однородное уравнение свободных поперечных колебаний ограниченной однородной струны 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 уравнение ( $n$ -dimensional wave equation)

●● волновое  $n$ -мерное неоднородное/однородное уравнение

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

●● волновое одномерное однородное уравнение, [PDE1046],

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

●● волновое одномерное неоднородное уравнение, [PDE1047],

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

●● волновое двумерное однородное уравнение, [PDE1048],

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

●● волновое двумерное неоднородное уравнение, [PDE1049],

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

Для реализации полноценного поиска необходимо использовать список ключевых слов и их английских эквивалентов. Приведем варианты ключевых слов и англоязычных эквивалентов для тезауруса по волновому уравнению в таблице 2. При составлении этого списка использовались различные источники, в частности ресурсы [11-14].

Таблица 2. Ключевые слова и англоязычные эквиваленты

Уравнение с частными производными (УЧП) Гиперболический тип уравнения Волновое уравнение	Partial differential equations (PDE) Hyperbolic partial differential equations Wave equation
--	--



Одномерное волновое уравнение	One-dimensional wave equation Wave equation 1-dimension 1-st order PDE First -order partial differential equation: Linear partial differential equations Quasilinear partial differential equations
Свободные колебания струны	Free string oscillation
Вынужденные колебания струны	Forced string oscillation
Однородное волновое уравнение	Homogeneous wave equation
Неоднородное волновое уравнение	Inhomogeneous wave equation
Колебания бесконечной струны	Infinite string oscillation
Колебания полубесконечной струны. Одномерный случай	Oscillations of an infinite string 1-D case. One-dimensional case
Малые поперечные колебания	small cross oscillations
Задача с начальными условиями. Коши задача  Решение задачи Коши	Problem with initial condition Cauchy problem <a href="#">1st order Cauchy problem</a> First order PDE Solution to the Cauchy problem
Даламбера метод. Метод бегущих волн. Лемма Даламбера	D'Alembert method. Wave propagation method. D'Alembert lemma.
Формула Даламбера. Решение Даламбера	D'Alembert formula. D'Alembert solution
Волна как передвижение отклонения по струне Распространение волн отклонения	Wave as a movement of a deviation on a string  Movement of waves of a deviation
Коэффициент одномерного волнового уравнения как скорость распространения волны	Coefficient of the one-dimensional wave equation as speed of distribution of a wave
Прямая волна. Обратная волна	Direct wave. Return wave
Передний фронт волны. Задний фронт волны	Forward front of a wave. Back front of a wave
Фазовая плоскость. Одиночная волна отклонения	Phase plane. Single wave of a deviation
Волны импульса. Отражение волн	Impulse waves. Reflection of waves
Распространение волн импульса	Movement of waves of an impulse
Незатухающие колебания	Not damped oscillations
Свободные колебания струны, закрепленной на обоих концах	Free fluctuations of the string fixed on both ends
Метод Фурье. Метод разделения переменных	Fourier's method. Fourier's method of separation of variables Separate variable method to a PDE Fourier's method of separation of variables to solve the boundary value problem
Гиперболические уравнения второго порядка. Телеграфное уравнение	Second Order Hyperbolic Equations Telegraph Equation
Параболическое уравнение второго	Second Order Parabolic Equations

порядка Трикоми уравнение	Tricomi Equation
------------------------------	------------------

Результаты представлены в виде массива, предназначенного для использования в системе GRAILS[15]. Используется онтологическая технология для описания предметной области обыкновенных дифференциальных уравнений и уравнений с частными производными (волновое уравнение). За основу принят тезаурус ОДУ[10] с переводом терминов на английские эквиваленты и результаты исследований авторов, в частности [16, 17].

## О "НОВЫХ" ТЕРМИНАХ

Для учета изменения свойств термина в процессе развития предметной области необходимо предложить механизм индексации терминов с учетом наследования первоначального значения (значений). Это непосредственно связано с проблемой наследования в онтологиях. Одним из механизмов может служить введение дополнительных признаков в статьи тезауруса, обозначающих связи термина с другими предметными областями (рис. 2).

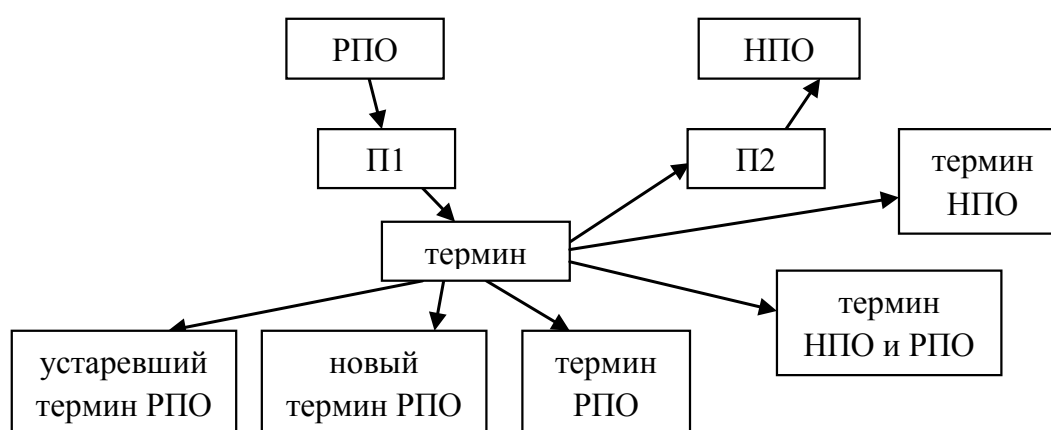


Рис.2. Технология онтологического подхода с учетом наследования

Здесь: "родовая" предметная область (РПО) термина - предметная область (ПО), где впервые употребили этот термин для определения какого-то понятия (П1); "новая" предметная область (НПО) - где этот термин стали употреблять для другого понятия (П2).

Показано, что свойства термина могут меняться в процессе расширения предметной области вплоть до его "отделения" в новую область. Если термин устаревает и используется для определения новых понятий, то можно говорить о появлении новой предметной области в коммуникационной среде. Для описания онтологий это важно, так как надо учитывать "новые" связи понятий и не учитывать "старые" в зависимости от поискового запроса, поскольку

добавление новых связей может привести как к уменьшению поискового шума, так и его увеличению.

## О "СТАРЫХ" ТЕРМИНАХ

Практика показывает, что накопленная в интернет информация может носить противоречивый характер, вплоть до отрицания старых понятий и заменой их новыми. Например, понятия "модель" и "моделирование" претерпевают изменения все время своего существования. Известны многочисленные примеры использования модных терминов, типа "нано" - термины, зачастую не имеющие никакого отношения к их первоначальному применению. Тем не менее, отвергать "модные" термины невозможно, так как они отражают текущее восприятие происходящего в науке и обществе.

При использовании интернета как основного справочного материала естественным образом не учитывается то, что осталось в старых монографиях, учебниках и статьях, которые по какой-то причине не попали в справочники сети. В этой связи возникает проблема, как не потерять накопленные знания и опыт предшественников, использовать их, чтобы не открывать заново колесо, но и не пользоваться ложной терминологией. Эта проблема появилась довольно давно, и многие библиотечные ресурсы пытаются преодолеть ее, создавая оцифрованные архивы научного наследия (в частности, есть известный проект с аналогичным названием [18]). Известно, что доступ через интернет к основным библиографическим базам появился уже в новое время, практически в 21 веке, хотя и базы эти и интернет уже были значительно раньше (оцифровка библиографических данных ведется различными коллективами с 70-х годов 20 в., а рождение интернета 60-е годы 20 в. ). Необходим был соответствующий сервис, технологии и правовое урегулирование. Теперь у нас есть возможность провести условную границу между библиографическими ресурсами до их цифровой эпохи и современными, в интернет. Это могут быть различные данные, но при научных исследованиях по какой-то предметной области их необходимо рассматривать в совокупности. Естественно проследить изменение терминологии в публикациях, так как это отражает новые результаты в науке и технике, например, взяв за основу "историческую" терминологию от происхождения термина в науке и до 21 в. и далее рассматривать "новую" версию в интернет (рис. 3.)

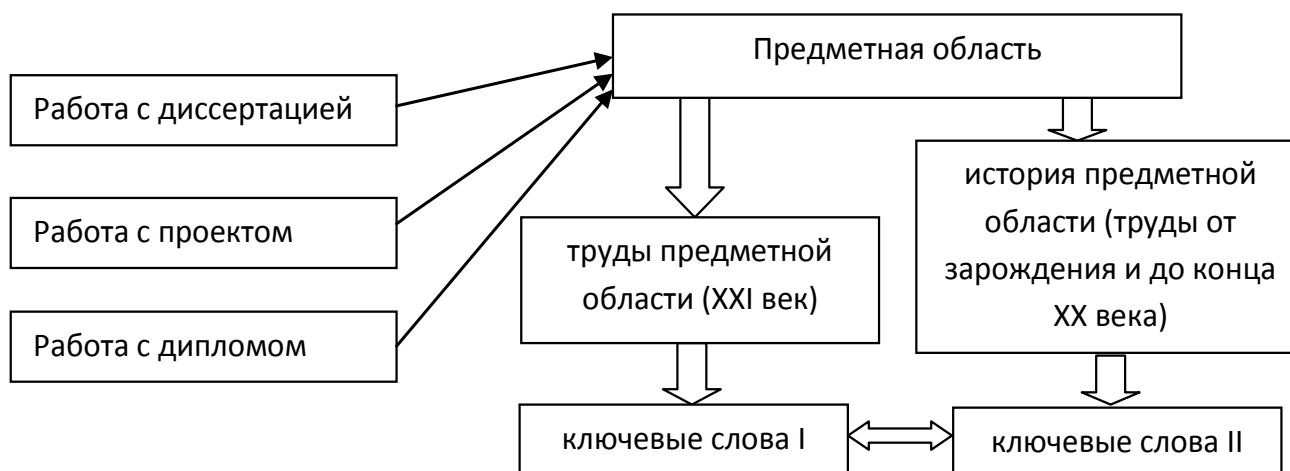


Рис.3 Схема использования «исторического» ресурса (вариант)

Множества ключевых слов (I и II) могут пересекаться, но в этом случае при поиске будет учитываться "первоисточник" и "происхождение" термина.

Можно проследивать не только историю развития предметной области, но и более достоверно судить о получении новых научных результатов, что немаловажно в контексте установления рейтингов ученых и их трудов.

Предложенные информационные ресурсы предполагается использовать как дополнение при создании онтологий предметных областей, в частности, "математика", и электронных библиотек, позволяющих поддерживать такие структуры данных.

## О РАЗВИТИИ ИНТЕГРИРОВАННОЙ СРЕДЫ

Данное исследование опирается на накопленный опыт участников проекта по теме исследований, примененный для создания прототипа "научной социальной среды" для предметных областей "волновое уравнение (ВУ)", "обыкновенные дифференциальные уравнения (ОДУ)", "уравнения в частных производных (УЧП)". Процесс интеграции данных, который наблюдается в интернет повсеместно, естественно захватил и научное сообщество. Это связано с необходимостью устранения дублирования, потребностью поддерживать актуальные данные и отслеживать недостоверную информацию. Возможность организовывать более эффективные научные коммуникации в сети приводит к следующей ступени развития цифрового научного сообщества, а именно системе с элементами социальной сети (например, СОЦИОНЕТ, LinkedIn и др.). Интеграция данных ученых РАН (ФАНО) также находится в стадии формирования связей и организации процесса коммуникации как внутри ресурса, так и с внешними ресурсами (рис. 4.).



Рис.4. Процесс интеграции данных на примере ВЦ РАН

Новые технологии представления библиографических данных с использованием онтологического подхода позволяют каждому автору (участнику научной сети) иметь свое представительство в предметной области (ПО на рис.5) в виде "тезауруса адресата" и общаться с коллегами, используя инструменты социальной сети, а также специальный набор сервисов для адресной доставки информации (подписка по ключевым словам (КС на рис. 5)).

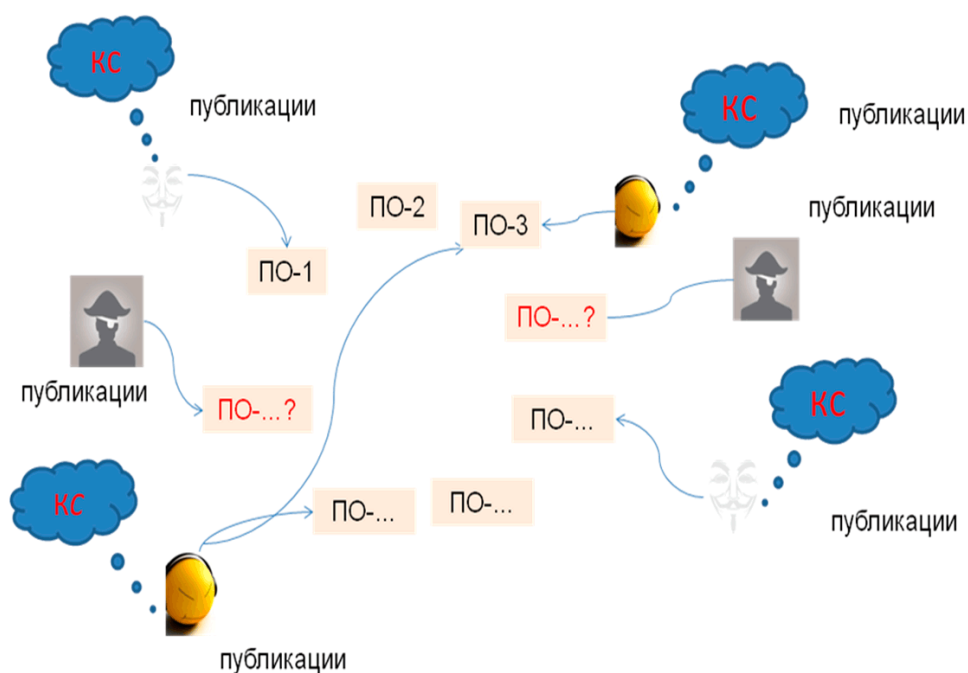


Рис.5. Схема научной социальной сети

## ЗАКЛЮЧЕНИЕ

В заключении приведем некоторые примеры из разрабатываемого информационного ресурса.

## Статья ЛСУ1

**Д:** Коши задача для волнового одномерного однородного уравнения свободных поперечных колебаний однородной бесконечной струны |PCPDE3515|

**Син:** Начальная задача о колебаниях однородной бесконечной струны |SYN8102(PCPDE3515)|

### Лит:

- И. Г. Араманович и В. И. Левин [A1], с. 24-51  
В. И. Арнольд [A4], с. 36-37  
В. М. Бабич, М. В. Капилевич, С. М. Михлин, Г. И. Натансон, П. М. Риз, М. М. Смирнов [B10], с. 38-43  
Л. Берс, Ф. Джон, М. Шехтер [B2], с. 16-21  
А. В. Бицадзе [B3], с. 160-172  
Т. Д. Вентцель, А. Ю. Горицкий, Т. О. Капустина, В. А. Кондратьев, Е. В. Радкевич, О. С. Розанова, Г. А. Чечкин, А. С. Шамаев, Т. А. Шапошникова [B10], с. 26-28  
В. С. Владимиров [B2], с. 221-240  
В. С. Владимиров, В. В. Жаринов [B9], с. 32-36, 167-187  
Д. М. Волков [B3], с. 80-91  
С. К. Годунов [Г4], с. 55-77  
Б. П. Демидович, В. П. Моденов [Д16], с. 201-208  
А. М. Колобов, Г. С. Неверов [К17], с. 61-71  
Н. С. Кошляков, Э. Б. Глинер, М. М. Смирнов [К3], с. 56-64, 68-69, 71-72  
В. П. Михайлов [M11], с. 266-283  
О. А. Олейник [O7], с. 199-215  
Ю. С. Очан [O2], с. 371-380  
И. Г. Петровский [П5], с. 105-111  
В. Н. Русак [P8], с. 75-79  
К. Б. Сабитов [C25], с. 82-90  
В. И. Смирнов [C16], с. 478-489  
С. Л. Соболев [C7], с. 44-47  
А. Н. Тихонов и А. А. Самарский [T3], с. 23-27, 49-59  
Ф. Трикоми [T2], с. 185-195  
Дж. Н. Шарма, К. Сингх [Ш7], с. 113-117

**МЗД:** Волновое одномерное однородное уравнение свободных поперечных колебаний однородной бесконечной струны |PDE1059(inf.str.)|<sup>6</sup>

$$u_{tt} = a^2 u_{xx}, \quad -\infty < x < \infty, \quad t > 0$$

$a^2 = T / \rho$ ,  $T$  – постоянная сила натяжения

$\rho$  – постоянная линейная плотность струны

Начальные условия

---

<sup>6</sup> inf. str. – бесконечная струна

$$U|_{t=0} = u(x,0) = f(x), \quad U_t|_{t=0} = \frac{\partial u}{\partial t}(x,0) = F(x), \quad -\infty < x < +\infty$$

**Род1:** Гиперболические задачи одномерные [PH1PDE4500]

**Род2:** Коши задача для волнового одномерного однородного уравнения [PCPDE4510]

**Ас1р:** Даламбера общее решение волнового одномерного однородного уравнения

$$u(x,t) = \Theta_1(x-at) + \Theta_2(x+at) \quad |\text{SPDE5007}|$$

**Ас2р:** Даламбера формула - решение Коши задачи для волнового одномерного однородного уравнения |\text{SPDE5002}|

**Ас3р:** Решение Коши задачи для волнового одномерного однородного уравнения свободных поперечных колебаний однородной бесконечной струны (ХМ) |\text{SPDE5009(ch.m.)}|

$$u(x,t) = \frac{f(x-at) + f(x+at)}{2} + \frac{1}{2a} \int_{x-at}^{x+at} F(s) dS, \quad -\infty < x < +\infty \quad (**)$$

**Ас4р:** Обобщенное решение Коши задачи для волнового одномерного однородного уравнения |\text{SPDE5010}|

**Ас5р:** Характеристик метод [MPDE4991]

**Прим:**

**Теорема** ([C25], 85-87). Если функции  $f(x) \in C^2(-\infty, +\infty)$ ,  $F(x) \in C^1(-\infty, +\infty)$ , то

существует единственное и устойчивое решение задачи Коши и решение определяется формулой (\*\*). См. также позиции «Д», «МзД», «Ас3р».

Формула (\*\*) определяет классическое решение Коши задачи [PCPDE3515]. Тем не менее эта формула может использоваться для представления обобщенного решения Коши задачи.

**Определение** ([K3], с. 71-72; [C25], с. 72-73; 87; [C7], с. 298-301; [B10], с. 18; [П5], с. 84-86)

Обобщенным решением Коши задачи для уравнения

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (***)$$

при начальных условиях  $u|_{t=0} = u(x,0) = f_0(x)$ ,  $u_t|_{t=0} = \frac{\partial u}{\partial t}(x,0) = F_0(x)$  называют

функцию  $u(x,t)$ , являющуюся пределом равномерно сходящейся последовательности решений  $u_n(x,t)$  уравнения (\*\*\*) при начальных условиях

$$u_n|_{t=0} = u_n(x,0) = f_n(x), \quad \left. \frac{\partial u_n}{\partial t} \right|_{t=0} = \frac{\partial u_n(x,0)}{\partial t} = F_n(x),$$

если последовательность функций  $f_n(x)$ , имеющих непрерывные вторые производные, сходится равномерно к  $f_0(x)$ , а последовательность функций  $F_n(x)$ , имеющих непрерывные первые производные, сходится равномерно к  $F_0(x)$ .

Имеем дескриптор:

«Обобщенное решение Коши задачи для волнового одномерного однородного уравнения» |SPDE5011|

Об интерпретации функций решения (\*\*\*) см. приведенную литературу.

**КлС:** однородная бесконечная струна, прямая волна, свободные поперечные колебания струны, обратная волна, |PrPDE3515|, |SPDE5009(ch.m.)|, |PCPDE0001|

УДК 517.956.32+517.955+517.958+534.112, бегущие волны, плоские волны.

**Замечание 11.** В позиции «КлС:» статьи ЛСУ1 поместили идентификаторы соответствующих дескрипторов по тематике статьи.

**Замечание 12.** По поводу решения уравнения (\*\*\*) см. краткую историю вопроса в ([Г4], с. 63-64; [С27], с. 171-173).

## **Статья ЛСУ2**

**Д:** Волновое одномерное однородное уравнение |PDE1046|

**Син1:** Уравнение малых колебаний с одной пространственной переменной однородное и линейное |SYN8103(PDE1046)|

**Син2:** Уравнение плоских волн |SYN8104(PDE1046)|

**Лит:** Большая часть приведенной литературы в настоящей работе, в том числе: [Г24], [К3], [Т3].

**МзД:**  $u_{tt} = a^2 u_{xx}, a > 0$

$a = \left\{ \sqrt{T/\rho}; \sqrt{E/\rho_0}; \frac{1}{\sqrt{LC}}; \sqrt{G \cdot J/K} \right\}$ , где

$T$  - постоянная сила натяжения;  $\rho_0$  - объемная плотность;

$\rho$  - линейная плотность;  $L$  - коэффициент самоиндукции;

$E$  - модуль Юнга;  $C$  - коэффициент емкости;

$G$  - модуль сдвига;  $J$  - полярный момент инерции;

$K$  - момент инерции единицы длины стержня

**Вид1:** Волновое одномерное однородное уравнение свободных поперечных колебаний однородной бесконечной струны |PDE1059(inf.str.)|

**Вид2:** Волновое одномерное однородное уравнение свободных поперечных колебаний однородной ограниченной струны с закрепленными концами |PDE1060(b. str.)|<sup>7</sup>

**Вид3:** Волновое одномерное однородное уравнение свободных продольных колебаний однородного ограниченного стержня постоянного сечения с закрепленными концами |PDE1061(b.rod)|

**Вид4:** Волновое одномерное однородное уравнение электрических колебаний в линии без искажений |PDE1062(el)|

---

<sup>7</sup> «b. str» -ограниченная струна; b.rod ограниченный стержень; el : электрические колебания



**Прим:** Соответствующие математические задачи для уравнений в позициях «Вид1» - «Вид4» не приводим.

«Вид4» - телеграфное уравнение в форме волнового уравнения (см. [К3], с. 95-97)

Переход от уравнения  $u_{tt} = a^2 u_{xx}$  к уравнению  $u_{tt} - u_{xx} = 0$  и обратно

осуществляются с помощью замены переменных, например (см. [Т3], с. 26):

$$u_{tt} = a^2 u_{xx} \rightarrow y = at \rightarrow u_{tt} - u_{xx} = 0$$

**КлС:** Волновое одномерное однородное уравнение |PDE1046| УДК 517.956.32+517.958

*Теоретические исследования ведутся в рамках темы "Модели управления колебательными процессами и градиентные модели сплошных сред и их терминологическое описание в информационной среде" (№ госрегистрации 01201352396) , практическая реализация поддерживается РФФИ (проект 13-07-00334).*

#### ЛИТЕРАТУРА

1. Roget P.M. Roget's thesaurus of English words and phrases. Longman, London, 1852, Penguin Books Ltd; Revised edition, 1998. 1140 p.
2. Enros P.C. May Dictionary and Thesaurus Projects, Scope and Status. Report Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, 1980.
3. Lewis A.C. Kenneth O. May and information retrieval in mathematics // Historia Mathematica V. 31, Issue 2, May 2004, P. 186–195.
4. Араманович И.Г., Левин В.И. Уравнения математической физики. Изд. 2-е, стереотипное. (Серия: Избранные главы высшей математики для инженеров и студентов втузов.) М.: Наука, 1969. 288 с.
5. Ильин В.А., Моисеев Е.И. Оптимизация граничного управления смещением на одном конце струны при свободном втором ее конце за произвольный достаточно большой промежуток времени // ДАН. Т. 417, N1, ноябрь 2007. С. 12-17.
6. Тимошенко С.П., Гудьер Дж. Теория упругости. М: Наука, 1979. 560 с.
7. Lurie S., Belov P., Tuchkova N. The Application of the multiscale models for description of the dispersed composites// Int. J. Computational Materials Science. A., 2004, 36(2):145-152.
8. Кошляков Н.С., Глинер Э.Б., Смирнов М.М. Дифференциальные уравнения математической физики. М.: Госуд. изд-во физ.-матем. лит., 1962. 767 с.
9. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Интернет и математические знания: представление уравнений математической физики в информационно-поисковой среде. М.:МАКС Пресс, 2008. 80 с.

10. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Тезаурус информационно-поисковый по предметной области: обыкновенные дифференциальные уравнения. М.:МАКС Пресс, 2005. 116 с.
11. Mathematics Stack Exchange URL: <http://math.stackexchange.com> (последнее обращение 28.05.2015)
12. Polyanin A. D., Zaitsev V. F. Handbook of nonlinear partial differential equations (2nd ed.). Boca Raton–London: Chapman & Hall/CRC Press. 2012. 1912 p. ISBN: 9781420087239
13. Lohwater's A. J. Russian–English Dictionary of the Mathematical Sciences. Second Edition, Revised and Expanded. Edited by R. P. Boas. American Mathematical Society. Providence, Rhode, Island, 1990. 343 с.
14. Math on Web <http://www.ams.org/mathweb/mi-classifications.html> (последнее обращение 28.05.2015)
15. The Grails Framework - Reference Documentation URL: <http://grails.org/doc/latest/guide/index.html>
16. Моисеев Е.И., Муромский А.А., Тучкова Н.П., Меденников А.М. Описание научных исследований и представления предметных областей в коммуникационной среде на основе контролируемой лексики // «Научный сервис в интернет: все грани параллелизма». Труды международной конференции», 23 - 28 сентября 2013 г. Новороссийск. М.: Изд-во МГУ, 2013. 589 с. (с. 520-521)
17. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Онтология научного пространства или как найти гения // Ж. Онтология проектирования. 2014. №4(14). С. 18-33.
18. Каленов Н.Е., Савин Г.И., Серебряков В.А., Сотников А.Н. Принципы построения и формирования электронной библиотеки "Научное наследие России"// Программные продукты и системы. 2012. Т.4(100). С. 30-40. (URL: <http://www.swsys.ru/index.php?page=article&id=3305>, обращение 04.09.2015)

# МУТАЦИОННОЕ ТЕСТИРОВАНИЕ СЕТЕВЫХ ПРОТОКОЛОВ С ИСПОЛЬЗОВАНИЕМ ФОРМАЛЬНЫХ МОДЕЛЕЙ

А.В. Никешин, Н.В. Пакулин, В.З. Шнитман

*Институт системного программирования Российской академии наук*

## 1 ВВЕДЕНИЕ

Тестирование является важной частью системы обеспечения корректности и надежности программных продуктов. Как правило, тестирование проводится на соответствие спецификации. На вход системе подаются данные предлагаемые спецификацией, затем поведение системы сравнивается с ожидаемым, и делается вывод о корректности работы системы. Однако, кроме воздействий определяемых спецификацией, необходимо проводить тестирование на входных данных, противоречащих спецификации или не определяемых ею. Такие ситуации постоянно возникают из-за ошибок пользователей или из-за целенаправленных действий злоумышленника. Особенно это актуально для систем безопасности.

В 1988 году Бартон Миллер (University of Wisconsin-Madison) предложил подавать на вход приложениям неструктурированные случайные данные. Данный метод, названный автором Fuzz-генератором, был опробован на системных утилитах ОС Unix и показал хорошие результаты: значительная часть приложений завершилась с фатальными ошибками или "зависли". Позже область тестирования была расширена на сетевые службы и приложения с графическим интерфейсом [1, 2]. Однако, весьма показательно, что применение этого метода для тестирования сетевых сервисов (за исключением некоторых клиентов) и X-серверов не дало результата: любая простейшая проверка входных данных, необходимая для серверных приложений, ограничивает применение данного метода. Дальнейшее развитие данного подхода привело к созданию более умных Fuzz-генераторов: на вход подаются более-менее структурированные данные, соответствующие конкретной тестируемой системе, в которых отдельные части или поля меняются или случайным образом, или в соответствии со схемой генератора тестов.

На данный момент существует большое количество различных генераторов некорректных данных (fuzz-генераторов) для разных классов тестируемых систем (файлов, драйверов, интернет-приложений, сетевых протоколов и др.). Нам, прежде всего, интересуют применение их для тестирования сетевых протоколов, и, особенно, протоколов безопасности.

Можно выделить два класса fuzz-генераторов:

- 1) самостоятельно создающие некорректные данные,
- 2) изменяющие (мутирующие) готовые корректные данные, которые

применяются при тестировании по принципам белого, серого или черного ящиков (white-, grey-, or black-box testing).

Кроме того, генераторы разделяют на "глупые" и "умные". Первые – не имеют никакого представления о структуре тестовых данных. Применительно к сетевым протоколам такие генераторы имеют довольно низкую эффективность. Использование в современных протоколах различных механизмов контроля данных (контрольные суммы, одноразовые значения, взаимосвязь различных сообщений и полей и т.п.) значительно сокращает область применения (а значит и класс обнаруживаемых уязвимостей) таких генераторов. "Умные" fuzz-генераторы используют знания о структуре изменяемых данных. Такая информация может быть получена с помощью анализа исходного кода, из формальной модели тестируемой системы или просто из внешнего файла, описывающего структуру данных.

При fuzz-тестировании на вход тестируемой системе подаются некорректные данные. Спецификация не всегда определяет поведение системы для таких случаев, поэтому часто ее невозможно использовать для оценки правильности поведения системы. Однако можно выделить множество состояний системы, которые являются нежелательными: завершение из-за фатальной ошибки, "подвисание", ошибки доступа к памяти. Для обнаружения таких ситуаций в основном и используется fuzz-тестирование. Для поиска ошибок доступа к памяти дополнительно используются анализаторы динамической памяти (memory debugger). Часто бывают случаи, когда некоторый модуль системы вызывает ошибку доступа к памяти, при этом сама система остается полностью работоспособной. Однако в сложных системах ошибки обращения к памяти часто используются злоумышленниками. Поэтому применение fuzz-тестирования является актуальным методом поиска проблем безопасности различных систем.

## **2. ТЕСТИРОВАНИЕ TLS С ИСПОЛЬЗОВАНИЕМ МУТАЦИЙ**

В нашей работе по тестированию протоколов безопасности мы используем технологию UniTESK, а для текущих экспериментов – разработанную нами ранее модель протокола TSL [3,4], поскольку она не только описывает сложную структуру данных, но и определяет отношения между различными ее элементами, а также внутреннее состояние и правила функционирования протокола.

Технология UniTESK предоставляет средства автоматизации тестирования соответствия формальным спецификациям [3]. В UniTESK тест представляет собой конечный автомат, с каждым переходом которого сопоставлено определенное тестовое воздействие. При выполнении перехода это воздействие передается на тестируемую реализацию, после чего регистрируются реакции реализации и автоматически выносятся вердикт о соответствии наблюдаемого поведения спецификации. Алгоритм обхода не

зависит от протокола и тестируемой системы, он реализован как компонент UniTESK.

Для тестирования протоколов, как правило, используются два сетевых узла: на одном узле установлена тестируемая реализация протокола, на другом, инструментальном узле, - модельная реализация под управлением UniTESK. Тестовыми воздействиями являются сообщения тестируемого протокола, сформированные модельной реализацией, а реакциями - сообщения от тестируемого узла.

Данный подход позволяет использовать для fuzz-тестирования как методы генерации, так и методы мутации данных. Методы генерации направлены на поиск потенциальных уязвимостей реализаций, выявленных при анализе спецификации протокола. Методы мутации используют в качестве входных данных правильные сообщения, сформированные модельной реализацией протокола. Использование готовой модели протокола TLS, позволяет мутировать данные на любом этапе обмена этого протокола. Стоит отметить, что в качестве исходных данных для мутации могут использоваться и некорректные сообщения, используемые, например, в методах генерации. Для непосредственного изменения данных используются механизмы, называемые операторами мутации. Операторы мутации определяют область значений и правила изменения конкретного типа данных.

Разработка велась с использованием инструмента автоматизированного тестирования JavaTesK [5]. Для мутационного тестирования протокола TLS были разработаны операторы мутации для некоторых основных типов данных, которые используются в формальной модели протокола.

Операторы мутации представляют собой механизмы внесения изменений в корректный поток обмена протокола. В работе [6] предложена классификация операторов мутации для протоколов по целевому воздействию:

- операторы, изменяющие порядок сообщений в потоке обмена,
- операторы, изменяющие порядок заголовков в конкретном сообщении,
- операторы, изменяющие значения полей заголовка.

В нашем случае изменение порядка сообщений реализовано встроенными средствами UniTESK и включено в тестовый набор протокола TLS, поскольку соблюдение порядка сообщений – одно из требований спецификации этого протокола. Заголовков (payloads) сообщения TLS не содержат. Поэтому мы ограничились третьим типом операторов. Эти операторы определяют область значений и правила изменения конкретного типа данных. Например, для целых типов данных определены следующие виды мутации:

- установление произвольного значения,
- установление значения нуль,
- установление максимального для данного типа значения.

Для байтовых массивов:

- изменение произвольного байта в массиве,
- изменение всех байтов в массиве,
- использование массива нулевой длины
- уменьшение/увеличение длины массива

Эти операторы являются универсальными для различных протоколов, поскольку структуры сообщений протоколов, как правило, представлены различными комбинациями двух типов данных: целых значений и байтовых массивов.

Кроме того, для некоторых типов данных протокола TLS определены свои операторы мутации. В основном, они являются расширением указанных выше операторов (например, оператор мутации для массива попарных байтов) и их использование ограничено данным протоколом.

Операторы мутаций задаются как классы на языке Java.

```
public class MutateOperator {
    public int Mutate(Object);
    public int MutateByteField(Byte, int);
    public int MutateIntegerField(Integer, int);
    public int MutateStringField(String, int);
    public int MutateByteArrayField(Byte[], int);
    public static MutateField findAnnotation(Field);
}
```

Спецификация тестового набора на JavaTESK задает структуры сообщений и данных протокола как классы на языке Java. Для того, чтобы указать генератору тестовых последовательностей, какие операторы можно применять к тем или иным полям сообщений, были введены специальные аннотации. В тексте модели TLS к полям модельных сообщений добавляются аннотации, задающие, какой оператор мутации может быть применен к данному конкретному полю.

```
@MutateField (mutate = true; comment = "")
public ProtocolVersion version;
```

Аннотация “@MutateField” выделяет поля модельных сообщений, которые требуется мутировать. Параметр “mutate” определяет, требуется ли применять мутацию в текущем тесте, поскольку структура некоторых сообщений может меняться в зависимости от согласованных между узлами параметров TLS-соединения. Параметр “comment” предназначен для дополнительной информации о типе изменяемых данных, что позволит выбрать необходимый оператор мутации.

Тестовый сценарий с использованием мутаций представляет собой смесь корректных тестовых методов и мутационных тестовых методов. Корректные тестовые методы предназначены для обхода достижимых состояний целевой системы. Благодаря особенностям алгоритма обходчика тестового автомата, в каждом достижимом состоянии будет вызван каждый тестовый метод, в том числе методы, подающие на вход системе искаженные сообщения. Такая комбинация тестовых методов позволяет с одной стороны, гарантировать, что fuzz-генератор сможет «преодолеть» все проверки в реализации и пройдет через значимые состояния, а с другой позволяет протестировать устойчивость реализации TLS к искаженным пакетам в каждом состоянии.

Данный вид тестирования опробован на нескольких реализациях серверов и клиентов протокола TLS, что позволило выявить значительное количество ошибок в реализациях.

В качестве дополнительного средства обнаружения ошибок реализаций мы используем анализаторы динамической памяти, что, с нашей точки зрения, может позволить обнаруживать возможные ошибки обращения к памяти, не видимые по внешнему поведению системы. Следует правда заметить, что использование анализаторов памяти возможно лишь для реализаций, к которым имеется полный доступ (в частности, не подходят реализации, установленные на сторонних интернет узлах).

### **3. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ РЕАЛИЗАЦИЙ TLS**

Отметим, что в процессе работы над проектом тестовый набор для реализаций TLS постоянно расширяется дополнительными возможностями [7]. В частности, для тестирования реализаций клиентов TLS был разработан практически новый тестовый набор, а также агент для взаимодействия с тестируемой реализацией. Соответственно расширяется и количество обнаруживаемых этими тестовыми наборами отклонений от спецификации и других ошибок реализаций. В качестве объектов тестирования мы использовали три реализации TLS-серверов и семь реализаций TLS-клиентов.

В качестве серверов для тестирования были выбраны следующие реализации:

1. почтовый сервер Postfix.2.9.3 с открытой реализацией протокола TLS openssl.1.0.1c под управлением операционной системы Red Hat Enterprise Linux 5.5 [8];
2. реализация TLS в виртуальной машине Java 1.7.0\_05 (Java Secure Socket Extension, JSSE) [9];
3. тестовый сервер TLS интернет ресурса <https://www.mikestoolbox.net>.

В реализации Postfix было выявлено 10 отклонений от спецификации, из них 2 с использованием некорректных данных. В реализации JSSE было

выявлено 10 отклонений от спецификации, из них 4 с использованием некорректных данных. В третьей реализации было выявлено 6 отклонений от спецификации, из них 3 с использованием некорректных данных.

В качестве клиентов были выбраны семь открытых реализаций:

Библиотечные реализации протокола:

- реализация TLS в виртуальной машине Java 1.7.0\_05 (Java Secure Socket Extension) (найдено 9 ошибок из них 5 с использованием некорректных данных);
- встроенная реализация клиента TLS библиотеки openssl-1.0.1j (найдено 5 ошибок из них 4 с использованием некорректных данных) [10].

Интернет браузеры:

- интернет браузер Mozilla Firefox 34.0.5 (найдено 4 ошибки из них 1 с использованием некорректных данных) [11];
- интернет браузер Opera 12.17 (найдено 3 ошибки из них 2 с использованием некорректных данных) [12];
- интернет браузер SRWare Iron 42.0.2250.1 [13] (найдено 2 ошибки, обе с использованием некорректных данных).

Почтовые клиенты:

- Mozilla Thunderbird 31.7.0 [14] (найдено 3 ошибки из них 2 с использованием некорректных данных)
- TheBat 6.8.2 [15] (найдена 1 ошибка)

## 4. ЗАКЛЮЧЕНИЕ

В данной работе представлен подход к разработке тестового набора для тестирования реализаций протокола TLS. Этот подход использует для автоматизации тестирования технологию UniTESK: тестовая последовательность строится динамически как обход некоторого автомата теста, для построения тестовых воздействий и вынесения вердикта о корректности наблюдаемого поведения реализации используется модель протокола.

Разработан метод расширения тестовых сценариев на основе моделей средствами мутационного тестирования. Такая комбинация тестовых методов позволяет с одной стороны, гарантировать, что fuzz-генератор сможет «преодолеть» все проверки в реализации и доберется до значимых состояний, а



с другой позволяет протестировать устойчивость реализации протокола к искаженным пакетам в каждом состоянии. Удалось совместно задавать спецификации поведения, спецификации тестовых сценариев и операторы мутации.

Таким образом, наш подход позволяет использовать для fuzz-тестирования как методы генерации, так и методы мутации данных. Благодаря тому, что JavaTESK основан на языке Java, удалось совместно задавать типы данных для представления сообщений, спецификации поведения, спецификации тестовых сценариев и операторы мутации. В результате нет семантического разрыва между моделью поведения, моделью тестирования и средствами мутации.

С нашей точки зрения, этот подход доказал свою эффективность, поскольку обеспечил обнаружение отклонений от спецификации и других ошибок во всех выбранных реализациях протокола.

Проект выполняется при поддержке РФФИ, проект № 13-07-00869 «Разработка метода оценки устойчивости протоколов безопасности к атакам».

## ЛИТЕРАТУРА

1. B. P. Miller, D. Koski, C. P. Lee, V. Maganty, R. Murthy, A. Natarajan, and J. Steidl. Fuzz revisited: A re-examination of the reliability of UNIX utilities and services // *Office*, vol. 1525, no. October 1995, pp. 1–23, 1995.
2. B. P. Miller, L. Fredriksen, and B. So. An empirical study of the reliability of UNIX utilities // *Commun. ACM*, vol. 33, pp. 32–44, December 1990.
3. Bourdonov I., Kossatchev A., Kuli Amin V., and Petrenko A. UniTesK Test Suite Architecture. // *Proceedings of FME 2002. LNCS 2391*, pp. 77-88, Springer-Verlag, 2002
4. Н.В.Пакулин, В.З.Шнитман, А.В. Никешин. "Разработка тестового набора для верификации реализаций протокола безопасности TLS // *Труды Института системного программирования РАН*, том 23, 2012 г. Стр. 387-404.
5. JavaTESK - <http://www.unitesk.ru/content/category/5/25/60/>
6. P. Tsankov, M. T. Dashti, D. Basin. SECFUZZ: Fuzz-testing SecurityProtocols // *Proceedings of the 7th International Workshop on Automation of Software Test (AST 2012)*, pp. 1-7, 2012.
7. Никешин А.В., Пакулин Н.В., Шнитман В.З. Автоматизация тестирования соответствия реализаций стандарту протокола безопасности транспортного уровня TLS // *Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление*. Выпуск 2(193)/2014. стр. 180-188. ISSN 2304-9766.
8. Postfix, <http://www.postfix.org>
9. Java™ Secure Socket Extension (JSSE), <http://docs.oracle.com/javase/6/docs/technotes/guides/security/jsse/JSSERefGuide.html>

10. OpenSSL Project, <https://www.openssl.org/>
11. Mozilla Firefox, <https://www.mozilla.org/ru/>
12. Opera, <http://www.opera.com/ru/>
13. SRWare Iron, [http://www.srware.net/ru/software\\_srware\\_iron.php](http://www.srware.net/ru/software_srware_iron.php)
14. Mozilla Thunderbird, <https://www.mozilla.org/ru/>
15. TheBat, <https://www.ritulabs.com/ru/products/thebat/>

# ОТКРЫТАЯ НАУКА

С.И. Паринов

*Центральный экономико-математический институт РАН*

**Аннотация.** В статье обсуждаются требования Открытой Науки, а также конфликт этих требований с коммуникационными возможностями традиционной системы научных публикаций. Рассматриваются возможности научных информационных систем для обеспечения научного сообщества средствами коммуникации в духе Открытой Науки. Представлены возможности системы Соционет, в которой реализованы все основные требования Открытой Науки.

## **Введение**

Данная статья продолжает серию моих публикаций, посвященных Открытой Науке. В первой статье из этой серии [6], опубликованной в 2009 г. было высказано мнение «... формируется комплекс условий для появления новой более эффективной организационной формы научно-исследовательской деятельности, которую мы назвали Открытой Наукой». В статье от 2010 года [7] мы представили концепцию виртуальной научной среды, предполагающую существенно более высокую степень открытости и прозрачности в деятельности ученых и научных организаций. С этого времени началось создание виртуальной среды "Открытая Наука", на базе научной информационной системы Соционет (<https://socionet.ru/>). Данная статья, третья по счету в этой серии, представляет краткую историю развития идей Открытой Науки в международном научном сообществе (раздел 1), текущее переосмысление требований к Открытой Науке (раздел 2), обсуждение препятствий для развития Открытой Науки, создаваемых традиционной системой научных журналов (раздел 3). В разделе 4 приведены данные из обзора действующих научных информационных систем. Раздел 5 посвящен системе Соционет и ее текущему вкладу в поддержку средств и методов Открытой Науки. В заключении подведены краткие итоги.

## **1. Краткая история Открытой Науки**

Термин Открытая Наука (Open Science) начал активно использоваться в 2008-2009 годах, как представляется, в результате осмысления научным сообществом важных изменений, происходящих в практике научной работы и, в частности, в научных коммуникациях, под влиянием новых информационных технологий.

Первоначально, это понималось как расширение идей и подходов, продвигаемых движением за программное обеспечение (ПО) с открытым кодом (Open Source Software), на другие аспекты научной деятельности [4]. Основной

контекст Открытой Науки в этом случае представлялся так, как это изображено на Рис. 1.

При этом цели Открытой Науки определялись следующим образом [1]:

- Прозрачность в методологии экспериментов, наблюдений и сбора данных;
- Публичная доступность и повторное использование научных данных;
- Публичная доступность и прозрачность научных коммуникаций;
- Использование онлайн-средств для поддержки научного сотрудничества.

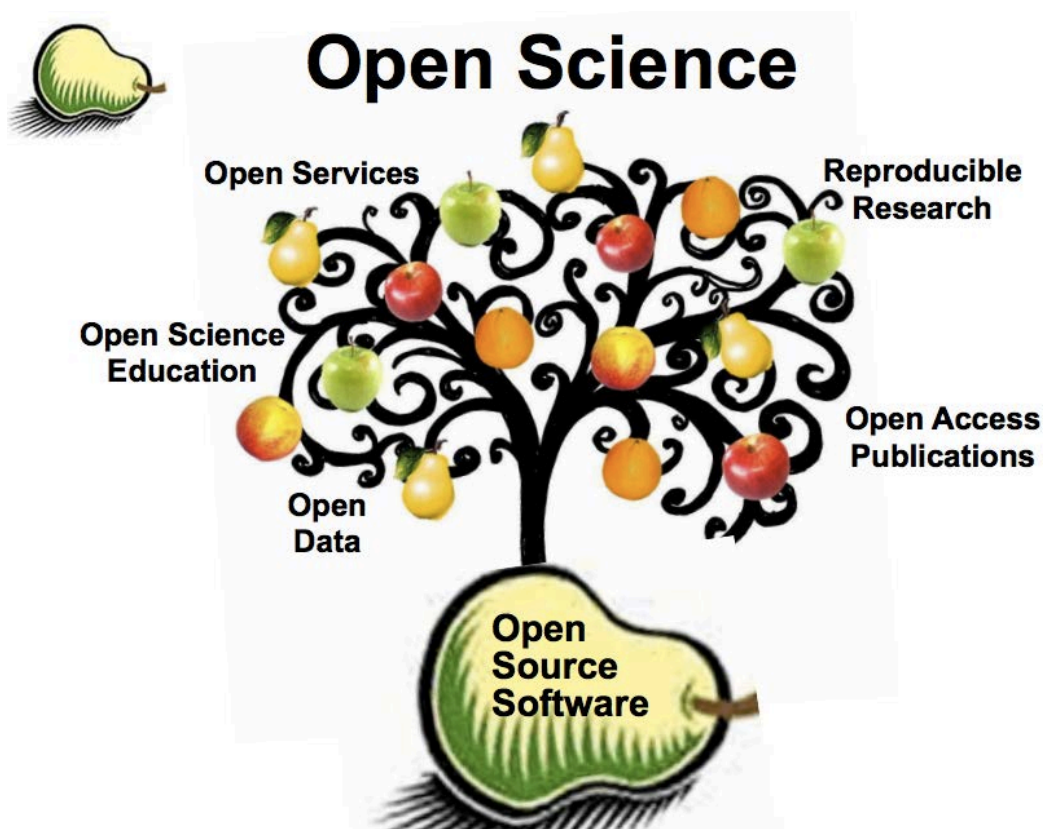


Рис. 1. Открытая Наука как развитие идей движения за ПО с открытым кодом.  
Источник - [2]

В последние годы контекст Открытой Науки существенно расширился и детализировался.

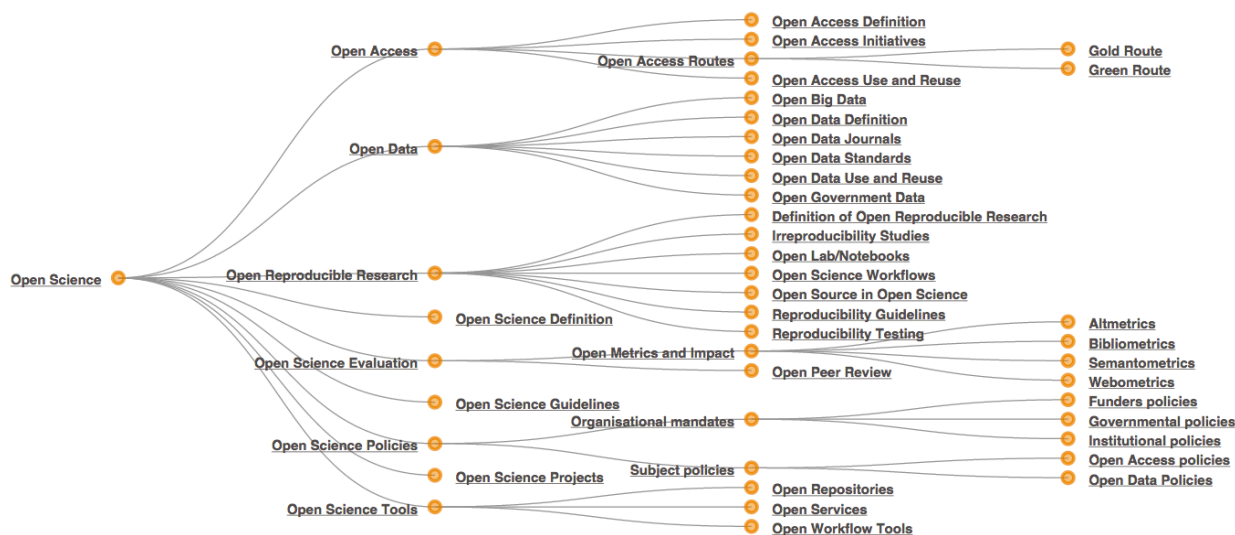


Рис. 2. Дерево понятий Открытой Науки. Источник – проект FOSTER.

Например, проект FOSTER<sup>8</sup>, целью которого является создание устойчивых механизмов для европейского научного сообщества по ускорению внедрения открытой науки в повседневную деятельность ученых, Открытую Науку представляет в виде довольно большого дерева понятий (и связанного с ним интерактивного глоссария), изображенного на рис. 2.

Участники проекта FOSTER определяют Открытую Науку следующим образом: проведение научных исследований таким образом, чтобы все желающие могли способствовать или участвовать в нем, при этом создаваемые исследовательские данные, лабораторные заметки и т.п. должны быть свободно доступны для использования, распространения и воспроизведения данного исследования<sup>9</sup>.

Это определение Открытой Науки, опубликованное недавно, видимо в 2014 г., по сути почти не отличается от приведенного выше определения 2009 года. Единственный новый момент – требование открытости процесса исследования для участия в нем посторонних исследователей.

Кроме естественного расширения содержания Открытой Науки, которое иллюстрируют рис. 1-2, усиливается критика традиционной научной системы с позиций Открытой Науки. Примером этого является доклад "Mapping the Future of Scholarly Publishing" [3], опубликованный в феврале 2015 г. международной рабочей группой «Инициатива Открытая Наука»<sup>10</sup>. Доклад содержит анализ с позиций Открытой Науки недостатков действующей системы научных публикаций и коммуникаций и предлагает некоторую программу действий. Среди принципов и направлений совершенствования

<sup>8</sup> <https://www.fosteropenscience.eu/>

<sup>9</sup> <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

<sup>10</sup> <http://nationalscience.org/portfolio/open-scholarship-initiative/>

системы научных публикаций и коммуникаций, в данном документе отмечаются следующие [3, стр. 5-6]:

- Научные работы должны быть свободно доступны, как только автор чувствует, что они готовы;
- Оценка работы должна начинаться с момента, когда она становится доступной, и продолжается пока это имеет смысл;
- Процесс оценки должен быть полностью открытым и каждый должен иметь возможность в этом участвовать;
- Оценка работы должна быть комплексной и включать такие факторы, как обоснованность, важность, направленность, этика и многие другие;
- Финансовая поддержка системы научных коммуникаций должна осуществляться организациями с гарантированной устойчивостью;
- Форма публикации и инфраструктура должны быть оптимизированы для поддержки разнообразного использования опубликованных работ

Эти формулировки полностью соответствуют идеям прозрачности и открытости научных исследований из ранее приведенных определений Открытой Науки. Однако в данном случае формулировки носят вид рекомендаций для ученых и научного сообщества по изменению традиционной формы научной деятельности.

## **2. Требования Открытой Науки**

Для уточнения деталей различных видов научной открытости, упоминаемых в трех выше приведенных источниках, рассмотрим их в контексте следующих укрупненных компонент научного процесса: (1) доступ ученого к существующему научному знанию; (2) оценка научной значимости и анализ применимости доступных ученому результатов исследований по отношению к своей работе; (3) использование ученым отобранных результатов исследований для производства нового научного знания; (4) коммуникации ученого с «соседями» по цепочкам глобальной научной кооперации; (5) оценка ученым своего текущего места в научной системе и возможных направлений для его улучшения.

Применительно к каждой из этих 5 компонент научного процесса требования Открытой Науки могут быть сформулированы в виде следующих требований:

**1. Открытый доступ к результатам исследований.** Это базовое и принципиальное требование, без выполнения которого не возможно все остальное. Содержание его достаточно хорошо известно, т.к. движение за Открытый доступ в науке имеет уже достаточно длинную историю (более 15 лет). Подборка публикаций на эту тему приведена в списке литературы [3, стр. 44].

**2. Открытая научная оценка результатов исследований.** Также достаточно известное требование, которое означает, что встретив публикацию заслуживающую внимания, ученый должен дать ей профессиональную оценку. Это требование реализовано в некоторых научных информационных системах в виде публичного (открытого) рецензирования<sup>11</sup>. Оценка, как правило, делается на основе заданного классификатора. Авторы оцененной публикации должны получать имейл-уведомление и система должна давать им возможности отреагировать на содержание оценки. Сделанные оценки, а также реакции на них авторов оцененных публикаций должны быть общедоступны на странице описания соответствующей публикации. Оценки также могут отображаться в данных о научной репутации как ученого, который их сделал, так и ученого, чьи работы были оценены.

**3. Открытое использование результатов исследований.** Новое требование. Технические возможности и примеры его реализации появились сравнительно недавно, в том числе, с развитием техники семантических связей в научных информационных системах [8]. Это требование означает, что ученый должен указать как или для чего им были использованы процитированные источники из списков литературы его публикаций. Выбор вариантов использования - на основе заданного классификатора. В комментариях ученый может дать пояснения. Система должна уведомлять авторов о всех случаях использования их результатов. В ответ авторы использованных результатов могут дать ценные советы и/или доработать свои результаты, чтобы пользователи этих результатов получили больший эффект. Подобные коммуникации между учеными должны содействовать возникновению долгосрочной научной кооперации между ними.

**4. Открытые научные коммуникации.** Требование не новое, но в нем появились новые моменты, вытекающие из предыдущих требований. Если ученый использует результат другого ученого, то он должен быть готов к ответной реакции в вариациях от протеста до предложения совместного развития идей. Содержание подобных коммуникаций - публично и доступно читателям публикаций, а также в профиле соответствующих авторов. Публичность содержания таких коммуникаций создает условия для конкуренции между учеными за право предоставить лучший результат для использования. Кроме этого система накапливает данные о подобных коммуникациях и обновляет статистические распределения показателей о том, как ученый реагирует на использование его результатов исследований, как другие ученые реагируют на использование их результатов исследований данным ученым, как ученый реагирует на реакции других ученых и как другие ученые реагируют на его реакции.

**5. Открытые данные о научной результативности и научной репутации ученого.** Предыдущие требования создают некоторые новые

---

<sup>11</sup> <https://peerj.com/>, <http://cyberleninka.ru/about>, <https://socionet.ru/> и др.

моменты для реализации этого вида открытости. Для каждого автора публикаций в научной информационной системе автоматически накапливается статистика, суммирующая различные характеристики его публикаций. Эта статистика включает: общее текущее количество публикаций, суммарное количество распознанных в системе цитирований публикаций автора, статистические распределения характера использования публикаций автора и т.п. Эти данные регулярно обновляются и характеризуют текущую результативность ученого не только с точки зрения выполненного им объема работ, но и демонстрируют, как научное сообщество использовало его результаты. В системе также автоматически собираются данные о научных оценках публикаций автора, о его реакции на эти оценки, а также о его оценках чужих публикаций и реакциях других ученых на оценки данного автора. Статистические распределения, построенные на основе этих данных, дают представление о том, какие категории оценок доминируют в настоящее время у определенного ученого по отношению к чужим работам и какова структура оценок его работ другими учеными. Авторы и все пользователи системы видят эти ежедневно обновляемые данные о своей научной репутации и могут на основе этого корректировать свои действия. Также они могут учитывать данные о научной репутации других ученых, например, если собираются с ними сотрудничать, или реагировать на их оценки по поводу своих публикаций.

В случае реализации этих 5 принципов открытости их совместное действие должно существенно изменить научный процесс. Эти изменения обещают разнообразные улучшения в работе научной системы. Однако на пути реализации этих требований стоят достаточно серьезные препятствия.

### **3. Традиционная система научных публикаций против Открытой Науки**

Уже упоминавшийся выше доклад "Mapping the Future of Scholarly Publishing" [3] также содержит достаточно подробный анализ недостатков традиционной системы научных публикаций. Один из выводов, который можно сделать из этого анализа - неспособность традиционной системы научных публикаций и коммуникаций, основанной, в первую очередь, на глобальной системе научных журналов, соответствовать требованиям Открытой Науки.

Проанализируем как выглядит процесс глобальных коммуникаций между авторами и пользователями опубликованных результатов исследований в том виде, как его поддерживает традиционная система научных журналов (рис. 3):

ученые получают доступ к журналам → читают статьи → выделяют в этих статьях важные для них фрагменты (научные артефакты) → формируют новое научное знание через создание как собственных новых артефактов, так и научных отношений между всеми связанными артефактами → превращают набор артефактов и отношений между ними в научную статью с цитатами, списком литературы и др. → добиваются публикации статей в журналах и полагаются на издателей в том, что потенциальные пользователи их



результатов исследований будут иметь доступ к этим журналам. Далее цикл повторяется.



Рис. 3. Научные коммуникации на основе традиционной издательской инфраструктуры

Про недостатки этого механизма коммуникации уже много написано (см. например, список литературы в [3, стр. 44]). Отметим еще некоторые новые важные моменты:

- потенциальные пользователи результатов исследований обычно сегментированы по доступу к системам агрегации журналов и автор, как правило, не может обеспечить всем им равный и немедленный доступ к его опубликованным результатам;
- при создании нового научного знания ученые мысленно оперируют научными артефактами и отношениями между ними, но для передачи этого знания другим ученым им приходится тратить силы и время на придание научному выводу формы статьи;
- при использовании (цитировании) результатов исследований действующая система не гарантирует что автор использованного результата узнает о всех случаях как, с какой целью и кем этот результат был использован, и тем более она не обеспечивает коммуникации между автором и пользователем, которые, если бы они были, могли бы улучшить качество результатов исследований обеих сторон (автора и пользователя), а также превратиться в продуктивную научную кооперацию данных ученых;
- реальное научное значение и влияние опубликованных результатов выясняется по итогам процесса «проб и ошибок», который возникает, когда ученые пытаются использовать эти результаты в своих исследованиях; качественные итоги этого процесса «проб и ошибок» не могут быть выражены просто наличием цитаты на соответствующий результат (они отсутствуют в индексе цитирования) и никак не связаны с импакт-фактором журнала, где этот результат опубликован.

С точки зрения современного уровня развития информационных технологий традиционная система научных журналов как способ глобальной коммуникации между авторами научных результатов и их пользователями выглядит устаревшей и неэффективной. Она и не может быть модернизирована в соответствии с требованиями Открытой Науки, т.к. эти недостатки заложены в ее конструкции.

Более эффективная система научных коммуникаций, создающая условия для применения принципов Открытой Науки, должна быть создана за рамками традиционной системы научных издательств.

В качестве одного из способов решений этой проблемы в упомянутом докладе [3] обосновывается необходимость создания единого научного репозитория (All-Scholarship Repository) для размещения исследователями своих научных выходов [3, стр. 35].

На наш взгляд, этого не достаточно. Технология открытых репозиториях достаточно хорошо реализует только первый из требований Открытой Науки: она обеспечивает открытый доступ к результатам исследований. Для реализации остальных принципов требуется более сложные информационные технологии. Здесь требуется научная информационная система или взаимосвязанный комплекс таких систем, которые предоставят ученым возможности создавать научные произведения, оперируя научными артефактами и отношениями между ними, а также будут включать сервисы, реализующие открытые коммуникации, аккумулирующие необходимую статистику и т.п.

#### **4. Научные информационные системы**

Текущая ситуация характеризуется большим количеством и разнообразием уже существующих научных информационных систем, в которых реализованы те или иные требования Открытой Науки. Хорошее представление об этом дает обзор «101 инновация в научных коммуникациях»<sup>12</sup>.

Данный обзор, к сожалению, не охватывает российские научные информационные системы<sup>13</sup>. Все перечисленные в нем научные информационные системы разбиты на следующие категории (рис. 4):

- Процесс научного открытия (Discovery)
- Научный анализ (Analysis)
- Написание научного произведения (Writing)
- Публикация научного произведения (Publication)
- Пропаганда и популяризация научных результатов (Outreach)

---

<sup>12</sup> 101 Innovations in Scholarly Communication, <https://innoscholcomm.silk.co/>

<sup>13</sup> Авторы анонсировали новую версию данного обзора, в котором уже более 400 систем - <http://bit.ly/innoscholcomm-list>

- Оценка научных результатов (Assessment)

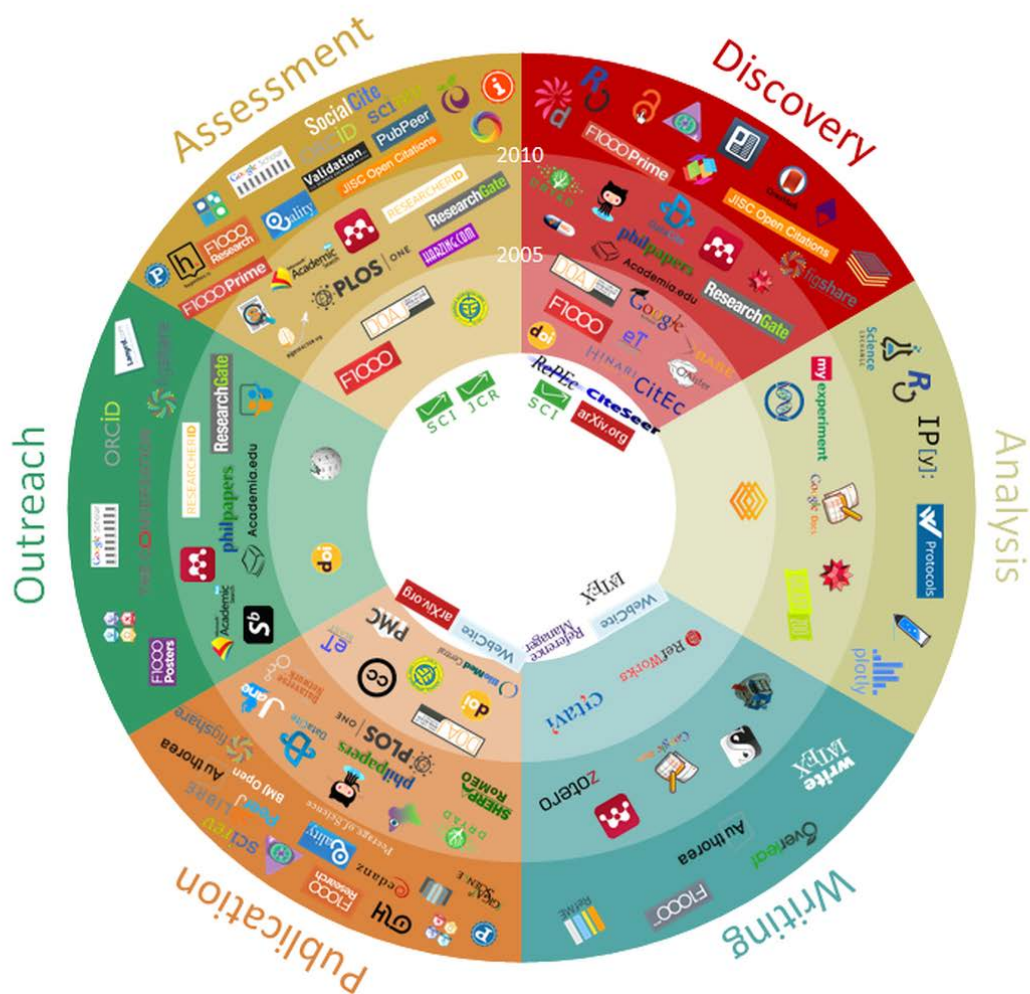


Рис. 4. 101 инновация в научных коммуникациях. Источник - <https://innoscholcomm.silk.co/>

Отнесение информационной системы к определенной категории сделано с определенной долей условности: авторы обзора сами определяли основную функциональность каждой системы.

Использованная здесь структура категорий отражает стадии научного процесса. Однако ее нельзя прямо соотнести с компонентами научного процесса, для которых мы выше рассматривали требования открытости.

По нашему мнению, ни одна из перечисленных в этом обзоре научных информационных систем не реализует все 5 требований Открытой Науки. Однако среди них можно выделить несколько наиболее популярных общедоступных систем: Academia.edu (<http://www.academia.edu/>), F1000 Research (<http://f1000.com/research>), Mendeley (<http://www.mendeley.com/>) и др.

К сожалению, в данный обзор не попала наша собственная система Соционет, которая реализует полный набор перечисленных требований открытости.

## 5. Научная информационная система Соционет

По состоянию на май 2015 г. общедоступная научная информационная система Соционет (<https://socionet.ru/>) - единственная, предлагающая пользователям функциональность, основанную на реализации всех 5 требований Открытой Науки:

1. авторы могут без ограничений самоархивировать свои научные выходы и выполнять таким образом требование открытого доступа к результатам исследований;
2. ученые могут выражать свое профессиональное мнение о доступных в Соционет материалов, заслуживающих внимание;
3. авторы могут для всех своих публикаций указать, в каком качестве ими были использованы (процитированы) источники из списка литературы;
4. авторы могут получать имейл-уведомления о всех случаях создания отношений (профессиональное мнение или использование) с их публикациями и публично реагировать на эти действия, при этом система Соционет информирует читателей о содержании этих открытых коммуникаций;
5. система Соционет собирает разнообразную статистику о действиях ученых и генерирует статистический портрет для каждого из них, которые включают данные о текущей научной результативности и научной репутации ученого.

Принципы организации и функционирования системы Соционет достаточно подробно описаны в публикации [5].

Одна из стратегических задач развития в Соционет технологий Открытой Науки – создание прототипа глобальной научной информационной системы для решения проблемы, создаваемой существующей научной издательской системой, которая как посредник между авторами результатов исследований и их пользователями плохо приспособлена для эффективных научных коммуникаций и формирования цепочек глобальной научной кооперации между исследователями, использующими результаты друг друга.

В Соционет созданы новые средства работы ученого, способствующие возникновению между авторами и пользователями результатов исследований прямых коммуникаций, сотрудничества и их более высокой ответственности за свои действия.

Практическое применение созданных в Соционет прямых средств сведения друг с другом «авторов» и «пользователей» результатов исследований в условиях конкуренции участников процесса научной кооперации создает ситуацию, при которой чем быстрее и с меньшими затратами найдут друг друга участники глобальной научной кооперации, тем эффективнее работает научная система.

Этот механизм функционирует следующим образом:

- Автор получает сигналы, когда кто-то использовал его результаты исследований;
- Автор может оказать помощь в правильном использовании его результатов исследований (кооперация);
- Автор может доработать свои результаты, чтобы усилить эффект от их использования (кооперация);
- Коммуникации между автором и пользователем формализованы и публичны (конкуренция);
- Все данные накапливаются и формируют публичный портрет ученого, включая его результативность и научную репутацию;

В этих условиях ответственность ученого за свои действия существенно повышается.

### **Заключение**

Одна из наиболее обсуждаемых проблем Открытого доступа и, как следствие, Открытой Науки – ослабление «входного» контроля качества публикаций, который выполняет рецензирование рукописей редколлегией журналов. Распространение открытого доступа путем самоархивирования учеными результатов своих исследований часто не предполагает рецензирование публикаций, а, значит, отсутствует их проверка на соответствие научным стандартам качества. Известно, что многие ученые не хотят тратить время и силы на чтение статей, если они заранее не знают уровень их качества. Разработанный в Соционет комплекс средств в определенной степени решает и эту проблему. Реализованные требования открытости порождают более жесткий самоконтроль и более высокую ответственность ученого за качество публикуемых материалов, а также за использование и оценку чужих результатов исследований и за свое поведение в профессиональных коммуникациях. С другой стороны, на страницах публикаций читателю доступны данные о том, кто и как их использовал, а также профессиональные мнения о них.

### **ЛИТЕРАТУРА**

1. Gezelter, D. What, exactly, is Open Science? Posted on July 28, 2009 by Dan Gezelter - <http://www.openscience.org/blog/?p=269>
2. Harris, N. Welcome to BOSC 2013. 14th Annual Bioinformatics Open Source Conference. July 19-20, 2013 Berlin, Germany - [http://www.open-bio.org/bosc2013/day1/BOSC2013\\_Intro\\_Nomi\\_Harris.pdf](http://www.open-bio.org/bosc2013/day1/BOSC2013_Intro_Nomi_Harris.pdf)
3. Open Science Initiative Working Group. Mapping the Future of Scholarly Publishing, 1st edition // Seattle: National Science Communication Institute, January 2015 - <http://nationalscience.org/wp-content/uploads/2015/02/OSI-report-Feb-2015.pdf>

4. Woelfle, M., Olliaro, P., Todd, M.H. Open science is a research accelerator // Nature Chemistry 3, 745–748 (2011).  
<http://www.nature.com/nchem/journal/v3/n10/full/nchem.1149.html>
5. Parinov S., Kogalovsky M., Lyapunov V. A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use // The Gray Journal (TGJ). An International Journal on Grey Literature. v. 10, i. 2, pp. 87-94, 2014 -  
<https://socionet.ru/publication.xml?h=RePEc:rus:isyigw:article-448>
6. Паринов С.И. Развитие электронных библиотек – путь к Открытой Науке // Труды 11й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009. -  
<https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:21>
7. Паринов С.И. Концепция виртуальной научной среды "Открытая Наука" // Научный сервис в сети Интернет, суперкомпьютерные центры и задачи: труды Международной суперкомпьютерной конференции (г. Новороссийск, 20-25 сентября 2010 г.). — М.: Изд-во МГУ, 2010. — С. 473-480. —  
URL: <https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:24>
8. Паринов С.И., Когаловский М.Р. Технология семантического структурирования контента научных электронных библиотек // Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011, Воронеж, Россия, 2011 -  
<https://socionet.ru/publication.xml?h=RePEc:rus:mqijxk:28>

# ГЕОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ АНАЛИЗА ГРАНИЦ КАРЬЕРОВ РУДНЫХ МЕСТОРОЖДЕНИЙ

Д.В. Петров, В.М. Михелев

*Белгородский государственный национальный исследовательский университет  
(НИУ «БелГУ»)*

Одним из популярных направлений в геоинформатике является решение задач анализа и моделирование геоданных с целью оптимизации стратегии поведения или использования имеющихся ресурсов.

К подобным задачам можно отнести:

- геостатистический анализ месторождений;
- планирование открытых горных работ, проектирование карьеров и массовых взрывов;
- горно-геометрический анализ и оптимизация границ карьера;
- визуализация результатов мониторинга технологических и природных процессов.

Фундамент современных геоинформационных систем закладывался в 80 года прошлого века, и они рассчитаны для работы на персональных ЭВМ. Однако, для решения комплексных задач освоения минеральных ресурсов необходимы мощные вычислительные средства совместной обработки больших массивов геологических, маркшейдерских, топографических, эколого-экономических и горно-технологических данных.

Основой для выполнения расчетов по оптимизации извлечения запасов является цифровая блочная модель месторождения. Вполне закономерно, что чем более многомасштабной и многокомпонентной является блочная (воксельная, ячеистая) модель рудного месторождения, тем более мощной, высокопроизводительной и надежной должна быть вычислительная система построения и динамического моделирования отработки запасов. В этой сфере наиболее эффективным становится использование передовых суперкомпьютерных технологий, реализующих распределенные вычисления на основе многопоточных алгоритмов и программ с использованием распараллеливания вычислений, таких, как технология MPI. В последние годы появилась также возможность задействовать для параллельных вычислений графические ускорители с программным ядром CUDA или OpenCL.

В этой связи становится актуальным решение задачи выделения некоторых частей геоинформационной системы в отдельный блок и реализации его в виде удаленного сервиса, предоставляющего клиентской части удобный интерфейс использования высокопроизводительных вычислительных комплексов.

В данной статье рассмотрены основные аспекты реализации информационной системы анализа моделей месторождений с целью нахождения оптимальных границ карьера для добычи полезных ископаемых открытым способом.

В современных геоинформационных системах анализа месторождений можно выделить несколько основных функциональных модулей (рис. 1):

1. подсистему ввода-вывода (импорта и экспорта данных)
2. подсистему прикладных функций (модуль поиска решения геоинформационной задачи)
3. подсистему визуализации данных
4. пользовательский интерфейс



Рис. 1. Основные компоненты геоинформационной системы

В обычном случае все компоненты реализуются в рамках одного приложения и работают на одной ЭВМ. Однако, т.к. блок поиска решения является особенно требовательным к вычислительным ресурсам, а остальные не обладают такой характеристикой его можно выделить в отдельный компонент. Это возможно, если в основу приложения положить клиент-серверную архитектуру. В такой системе клиента берет на себя функции импорта и экспорта данных, визуализации и организации взаимодействия с пользователем, а сервер реализует функции получения задачи клиента, их решение и отправку решения обратно клиенту (рис. 2).

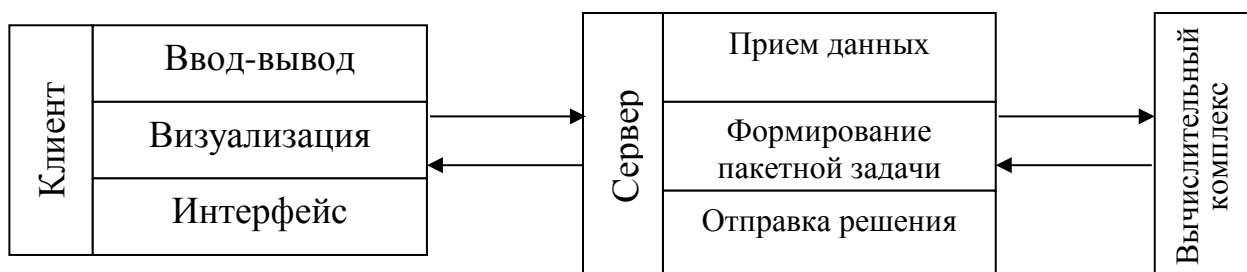


Рис. 2. Структура ГИС, основанной на клиент-серверной архитектуре



Система, построенная таким образом, позволяет использовать в качестве платформы для проведения сложных расчетов практически любые высокопроизводительные вычислительные ресурсы.

Описанный подход был применен для реализации информационной системы для анализа месторождений с целью поиска предельных границ карьера. Данная задача является одним из ключевых этапов проектирования открытой разработки недр. Ее решение позволяет оценить предельные значения получаемой прибыли от разработки месторождения при текущей цене на полезные компоненты. Также она является основой для дальнейших этапов проектирования таких, как прокладка кратчайших путей съездов в горных выработках и определение оптимальных мест расположения отвалов пустых пород.

В качестве вычислительного алгоритма используется параллельный генетический алгоритм поиска предельных границ карьеров, описанный в работах [1,2]. Реализация данного алгоритма использует несколько уровней параллелизма посредством выделения в структуре вычислений нескольких уровней, каждый из которых работает на различных технологиях параллельного программирования (MPI, OpenMP, CUDA) образуя одну из распространенных связок, например, MPI+OpenMP или MPI+CUDA[3].

Такой алгоритм хорошо накладывается на архитектуру больших гетерогенных распределенных вычислительных систем и позволяет равномерно разнести нагрузку по вычислительной системе, максимально эффективно используя многоядерные и гибридные вычислительные узлы.

Реализованная система состоит из следующих компонентов:

1. клиентское приложение
2. сервер обработки задач
3. вычислительный кластер

Клиентское приложение позволяет загружать модель месторождения на сервер и визуализировать результаты моделирования. Оно не производит расчетов, вместо этого отправляет запрос с параметрами задачи на расчет на сервер. Сервер анализирует запрос, оценивает примерное время решения задачи, определяет необходимое количество вычислительных ресурсов, формирует пакетную задачу на расчет для менеджера задач вычислительного кластера и ставит ее в очередь. Когда на кластере появляется необходимое количество свободных вычислительных узлов, задача поступает на выполнение. Об окончании расчетов менеджер задач кластера уведомляет сервер посредством стандартного механизма нотификации (посредством отправки письма на специальный email). После этого сервер получает результаты расчетов с кластера и отправляет их клиенту. Общая схема взаимодействия компонентов системы показана на рис. 3.

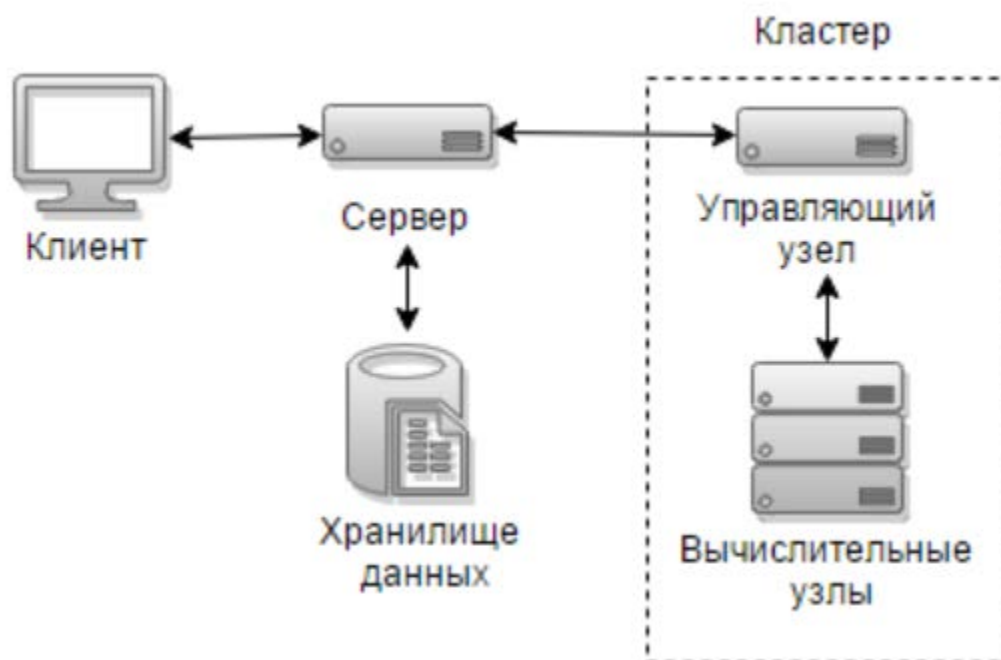


Рис. 3. Схема взаимодействия компонентов информационной системы

В качестве вычислительного комплекса использовался суперкомпьютер Белгородского государственного национального исследовательского университета. Основные характеристики кластера приведены в таблице 1.

Таблица 1. Суммарные технические характеристики кластерной системы

Характеристика	Значение
Семейство процессора	Intel Xeon
Частота процессора	2.4ГГц
Количество процессоров	40
Количество ядер	320
Объем ОЗУ	1280Гбайт
Объем HDD	8 Тбайт
Сеть	10 Гбит/сек.

Работа системы тестировалась на нескольких блочных моделях месторождений с различными типами распределения полезных компонентов. Модель со случайным равномерным распределением генерировалась специально для тестов. Модель с ярко выраженным рудным телом создана на основе результатов моделирования и подсчета запасов Жайремского месторождения в Казахстане, опубликованных в работах [4, 5]. Результаты моделирования приведены на рис. 4.

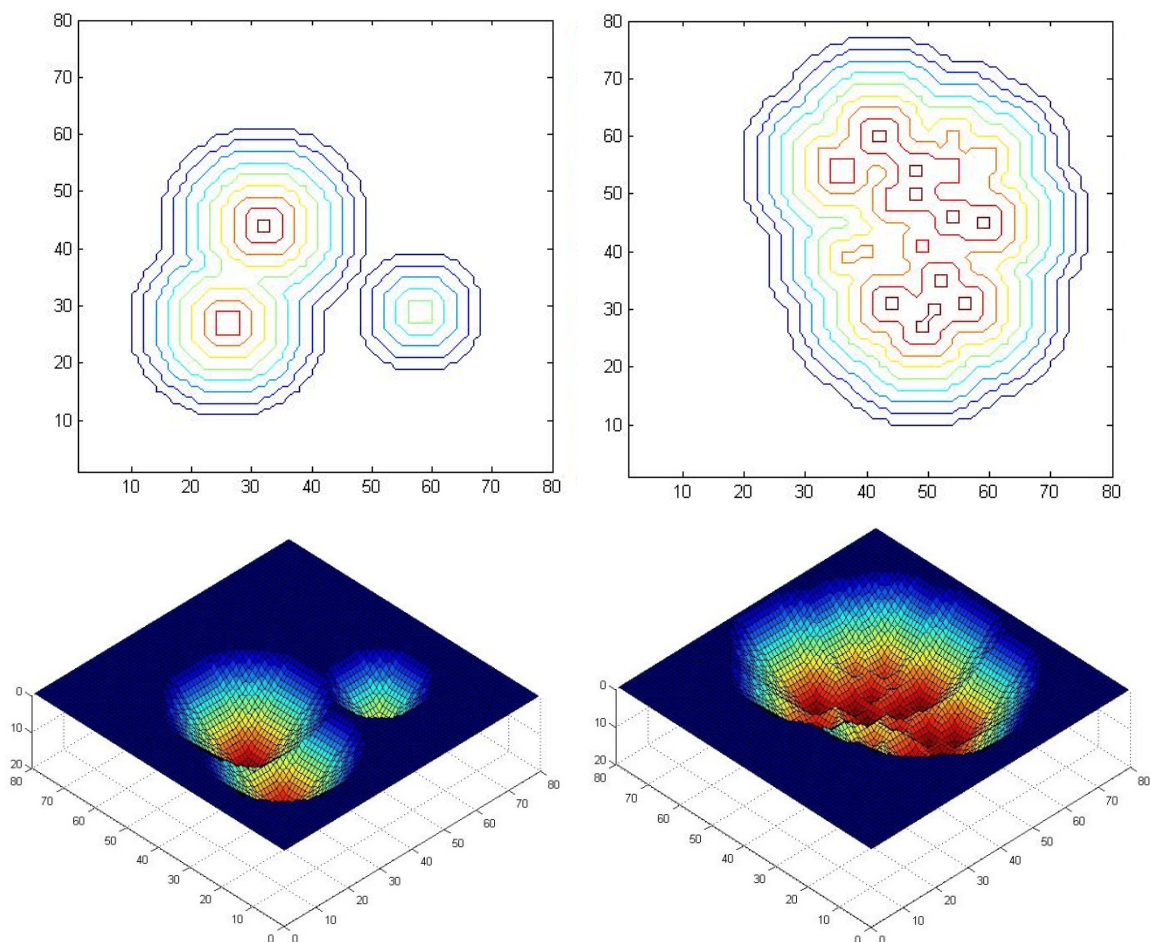


Рис. 4. Изображения карьеров, полученных в результате анализа месторождений  
 (а) со случайным распределением полезных компонентов и  
 (б) с ярко выраженным рудным телом

Геоинформационная система, построенная на принципах, описанных в статье обладает рядом преимуществ:

1. сокращается время поиска решения за счет применения параллельных алгоритмов и высокопроизводительного вычислительного оборудования;
2. один сервер позволяет обслуживать несколько клиентских станций за счет организации очереди задач или использования нескольких вычислителей одновременно;
3. повышается надежность хранения данных за счет их централизованного хранения и упрощения процесса резервного копирования;
4. появляется возможность строить приложения, использующие вычислительные ресурсы как сервис (пользователь имеет возможность приобрести не только программную систему, но и вычислительные ресурсы на удаленном вычислительном комплексе для решения своих задач).

## ЛИТЕРАТУРА

1. Петров Д.В., Михелев В.М. «Моделирование карьеров рудных месторождений на высокопроизводительных гибридных вычислительных системах», Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2014. Т. 3. № 3. С. 124-129.
2. Михелев В.М., Васильев П.В., Петров Д.В. «Суперкомпьютеры, как средства моделирования граничных контуров карьеров рудных месторождений», Вопросы радиоэлектроники. Серия "Электронная вычислительная техника (ЭВТ)" Выпуск 1, Москва 2013., с. 5-10
3. Стронгин Р. Г., Гергель В. П., Гришагин В. А., Баркалов К. А. Параллельные вычисления в задачах глобальной оптимизации: Монография / Предисл.: В. А. Садовничий. – М.: Издательство Московского университета, 2013. – 280 с., илл.
4. Селифонов С.Е. Агафонов В.А., Моргунова Т.В., Васильев П.В., Буянов Е.В. Компьютерная технология подсчета промышленных запасов рудных месторождений с использованием программы GEOBLOCK // Минеральные ресурсы Казахстана. 2000. т.12. №13. С.58-62
5. Васильев П.В., Буянов Е.В. О методике совместной работы программ MapInfo и Geoblock по оконтуриванию и подсчету запасов рудных месторождений // Информационный Бюллетень ГИС Ассоциации. 2000. №2. С.32-33

# СРАВНЕНИЕ ТРЕХ ПРОЕКТОВ ВЕБОМЕТРИЧЕСКОГО РАНЖИРОВАНИЯ НАУЧНЫХ ОРГАНИЗАЦИЙ

А.А. Печников

*Институт прикладных математических исследований КарНЦ РАН*

## **Введение**

В 2014 году группой по вебометрике Института прикладных математических исследований КарНЦ РАН был завершён проект «Вебометрический рейтинг научных учреждений России» (грант РГНФ №12-03-12001) [1]. Основными результатами проекта для автора стали не только (и не столько) разработанная методика ранжирования, способы сбора данных, исправление ошибок измерений и собственно результаты ранжирования, сколько «концептуальные», если так можно выразиться, вопросы, постоянно встававшие в процессе работы. Один из вопросов заключается в том, насколько близки (в каком-либо смысле) результаты ранжирования веб-ресурсов одних и тех же организаций в различных проектах.

Здесь мы попробуем сравнить три проекта:

1. «Ranking Web of Research Centers» [2], посвященный ранжированию научных организаций мира – один из подпроектов большого проекта испанской группы Cybermetrics Lab. Опубликованы данные ранжирования за январь 2015 года; около 8000 научных организаций, в том числе 182 российских. Будем называть его Sub-Lab.

2. «Рейтинг сайтов научных учреждений СО РАН», Института вычислительных технологий СО РАН [3]; ранжирование за сентябрь 2014 года, количество организаций – 91. Будем называть его СО-ИВТ.

3. «Вебометрический рейтинг научных учреждений России»; 397 научных учреждений РАН; ранжирование за декабрь 2014 года; назовём его РАН-ИПМИ.

Возможностей для сравнения проектов СО-ИВТ и РАН-ИПМИ было бы достаточно, т.к. пересечение целевых множеств содержит 82 организации, а для сравнения трёх проектов дело оказалось немного хуже, всего 18 организаций – но все же лучше, чем ничего.

## **Об индикаторах и средствах измерения**

Основными вебометрическими индикаторами сайта (по терминологии Cybermetrics Lab) являются следующие:

1. размер сайта ( $S$  – size) – общее количество страниц;
2. видимость сайта ( $V$  – visibility) – количество уникальных гипертекстовых ссылок на сайт с других веб-ресурсов;

3. количество полнотекстовых файлов на сайте ( $R$  – «rich files») – суммарное количество файлов с расширениями PDF, DOC, PS, PPT и т.д.;

4. научность сайта ( $Sc$  – «scholar»).

Изначально принималось за правило, чтобы значения индикаторов можно было измерить с помощью известных поисковых систем. В CybLab используются Google, Yahoo, Live Search и Exalead, в СО-ИВТ – Яндекс, Google и Bing, в РАН-ИПМИ – Яндекс и Google.  $Sc$  везде измеряется Google Scholar.

В проекте [1] значение  $Sc$  для конкретного сайта мы измеряли с помощью запроса к Google Scholar (<https://scholar.google.ru>) вида «site:доменное имя сайта», набираемого в поисковой строке.

Результаты сравнения:

- в каждом проекте мы имеем один и тот же базовый набор индикаторов;
- в СО-ИВТ он дополнен тИЦ Яндекса [4];
- в РАН-ИМПМИ он дополнен «внутренней ссылочной популярностью» (см. <http://webometrics-net.ru/section.php?id=38>), измеряемой VeeCrawler [5];
- средства измерения у всех трёх проектов различные, но имеются и общие;
- CybLab не использует Яндекс.

### **О целевых множествах**

В CybLab и СО-ИВТ целевое множество ранжирования фактически определяется как доменные имена официальных сайтов. В проекте РАН-ИМПМИ единицей анализа называется множество веб-сайтов информационного веб-пространства научного учреждения РАН, рассматриваемых как единое целое, идентифицируемое по названию научного учреждения, и значения индикаторов единиц анализа вычисляются как сумма значений индикаторов для всех сайтов веб-пространства организации.

Результаты сравнения:

- принципы формирования целевых множеств проектов CybLab и СО-ИВТ одинаковые;
- принципы формирования целевого множества проекта РАН-ИМПМИ существенно отличаются.

### **Об ошибках измерений**

В проекте РАН-ИПМИ возникали ситуации, которые рассматриваются как ошибки поисковых систем, когда, к примеру, при измерениях  $S$  с помощью Яндекса и Google получаемое количество страниц выражалось сотнями тысяч для небольших сайтов. Очевидно, что во всех трех проектах сбор значений индикаторов выполняется в автоматизированном режиме с помощью программ, подобных скраперу РАН-ИПМИ [6], поэтому вопрос об ошибках измерений актуален.

В проекте CybLab и СО-ИВТ внимание на этом вопросе не акцентируется. В проекте РАН-ИПМИ предложен подход «сглаживания» ошибок [7]. Статья имеется в открытом доступе на E-library.ru по адресу <http://elibrary.ru/download/93188149.pdf>.

Результаты сравнения:

- сравнение процедур «сглаживания» ошибок измерений значений вебметрических индикаторов не представляется возможным.

### **О ранжировании**

Процедура ранжирования в CybLab описана неясно, в [8] говорится о коэффициентах при  $V$ ,  $S$  и  $R$ , но индикатор  $Sc$  ещё не упоминался. Значения «одинаковых» индикаторов, полученных разными поисковыми системами, вначале усредняются. По-видимому, ранжирование проводится на основании значений функции вида

$$W = k_V * V_r + k_S * S_r + k_R * R_r + k_{Sc} * Sc_r,$$

где  $V_r$ ,  $S_r$ ,  $R_r$ ,  $Sc_r$  – значения рангов по соответствующим индикаторам, а  $k_V$ ,  $k_S$ ,  $k_R$ ,  $k_{Sc}$  – весовые коэффициенты. Можно считать, что наибольшая значимость придается  $V$  (половина от общей значимости), затем следуют  $S$ ,  $R$  и  $Sc$ .

В СО-ИВТ «обобщенные» значения  $S$ ,  $V$  и  $R$  усредняются, далее вычисляются ранги  $S_r$ ,  $V_r$  и  $R_r$ . Для  $Sc_r$ : сперва получают по отдельности ранги для  $Sc$  по Яндекс и Google, затем эти ранги суммируются, а уже потом организация, у которой полученная сумма наименьшая, получает ранг  $Sc_r = 1$ , и т.д. Для каждой организации вычисляется  $W = V_r + S_r + R_r + Sc_r$ . Итоговый ранг (позиция в текущем рейтинге), будет тем выше, чем меньше значение  $W$ .

В РАН-ИПМИ ранжирование единиц анализа научных учреждений рассматривается как задача о групповом выборе, для решения которой используется правило Борда [9].

Результаты сравнения:

- в каждом из проектов используются собственные процедуры ранжирования;
- в проектах CybLab и СО-ИВТ есть некоторое сходство.

### **Сравнение результатов ранжирования на примере Сибирского отделения РАН**

Пересечение целевых множеств проектов CybLab, СО-ИВТ и РАН-ИПМИ дает нам множество, содержащее доменные имена 19 научных организаций СО РАН, приведенных в таблице 1.

**Таблица 1. 19 научных организаций Сибирского отделения РАН**

<b>№</b>	<b>Название учреждения РАН</b>	<b>Адрес сайта</b>	<b>SybLab</b>	<b>СО-ИВТ</b>	<b>РАН-ИПМИ</b>
1	Сибирское отделение РАН	sbras.ru	1	1	10
2	Институт цитологии и генетики СО РАН	bionet.nsc.ru	11	6	4

3	Институт вычислительных технологий СО РАН	ict.nsc.ru	21	5	1
4	Институт систем информатики имени А.П. Ершова СО РАН	iis.nsk.su	22	10	7
5	Институт вычислительной математики и математической геофизики СО РАН	sscc.ru	25	7	8
6	Институт солнечно-земной физики СО РАН	ru.iszf.irk.ru	33	16	20
7	Институт катализа им. Г.К.Борескова СО РАН	catalysis.ru	44	12	15
8	Институт систем энергетики им. Л.А. Мелентьева СО РАН	sei.irk.ru	51	19	25
9	Институт сильноточной электроники СО РАН	hcei.tsc.ru	53	23	9
10	Институт физики им. Л.В. Киренского СО РАН	kirensky.ru	66	11	14
11	Институт гидродинамики им. М.А. Лаврентьева СО РАН	hydro.nsc.ru	69	37	31
12	Институт химической кинетики и горения СО РАН	kinetics.nsc.ru	76	29	24
13	Институт физики полупроводников им. А.В. Ржанова СО РАН	isp.nsc.ru	77	30	38
14	Институт автоматики и электрометрии СО РАН	iae.nsk.su	91	20	32
15	Институт физики прочности и материаловедения СО РАН	ispms.ru	99	31	26
16	Институт вычислительного моделирования СО РАН	icm.krasn.ru	104	24	22
17	Институт космических исследований и аэронавтики им. Ю.Г.Шафера СО РАН	ikfia.ysn.ru	113	56	37
18	Институт химической биологии и фундаментальной медицины СО РАН	niboch.nsc.ru	114	45	35
19	Институт "Международный томографический центр" СО РАН	tomo.nsc.ru	149	54	47

Для значений из таблицы 1 вычисленные значения коэффициента ранговой корреляции Кендалла [10]  $\tau$  при уровне значимости  $\alpha=0.05$  попарно равны соответственно:

$$\tau(\text{SybLab, СО-ИВТ})=0.77, \tau(\text{SybLab, РАН-ИПМИ})=0.67, \\ \tau(\text{СО-ИВТ, РАН-ИПМИ})=0.74.$$



Коэффициент ранговой конкордации Кендалла-Смита  $W=0.92$ .

### **Заключение**

Несмотря на различия в подходах по трём проектам (которые в ряде случаев представляются достаточно существенными), значения коэффициента ранговой корреляции, вычисленные на «пересекающемся» целевом множестве, показывают значимую связь. Более того, коэффициент ранговой конкордации говорит о наличии высокой степени согласованности, если рассматривать итоги ранжирования по каждому проекту как мнения проектов-«экспертов».

Почему значения коэффициентов  $\tau$  и  $W$  настолько большие? У автора есть на этот счёт ряд объяснений, выходящих по объёму за рамки расширенных тезисов.

Одно из них заключается в том, что множество из 19 сайтов, полученное как пересечение целевых множеств проектов CybLab, СО-ИВТ и РАН-ИПМИ, представляет собой множество «хороших» сайтов с точки зрения их сканирования краулерами поисковых систем, используемых для измерения индикаторов. Это подтверждается нашей практикой сканирования сайтов с помощью VeeCrawler: структура сайтов прозрачна, отсутствуют т.н. «паучьи ловушки» типа бесконечных календарей и т.д. Отсюда можно говорить о том, что результаты измерений одного и того же индикатора с помощью одной и той же поисковой системы с несколько разными синтаксисами поисковых запросов вполне могут давать сильно коррелирующие значения.

Что дают полученные результаты конкретному научному учреждению и директивному органу?! – В своих ответах автор исходит из того, что мы занимаемся ранжированием сайтов не для того, чтобы назвать лучших и худших (или не только для этого).

Поэтому для научного учреждения можно использовать результаты любого проекта из трёх (а лучше всех трёх), лишь бы его сайт был в списке целевого множества. Данные проектов предоставляют возможность заинтересованным лицам хотя бы чуть-чуть проанализировать ситуацию и попробовать ответить на вопрос типа: почему сайт моего института имеет рейтинг ниже, чем сайт института, где работает мой коллега (директор, заведующий лабораторией, научный сотрудник, веб-мастер и т.д.)? И сделать свой сайт ещё лучше.

Для директивных органов, например, Федерального агентства научных организаций (если оно займется вплотную веб-сайтами институтов, создаваемых, вообще-то, за бюджетные деньги), следует рекомендовать в качестве основы любой из российских проектов: обязательное использование данных Яндекса, более полный список организаций, понятные принципы, положенные в основу ранжирования.

## ЛИТЕРАТУРА

1. Вебметрический рейтинг научных учреждений России. URL: <http://webometrics-net.ru>.
2. Ranking Web of Research Centers. URL: <http://research.webometrics.info>.
3. Рейтинг сайтов научных учреждений СО РАН. URL: <http://www.ict.nsc.ru/ranking>.
4. Что такое ТИЦ. URL: <http://help.yandex.ru/catalogue/citation-index/tic-about.xml>.
5. Pechnikov A.A., Chernobrovkin D.I. Adaptive Crawler for External Hyperlinks Search and Acquisition // Automation and Remote Control, 2014, Vol. 75, No. 3, P. 587–593.
6. Нвохири А.М., Огийко А.А., Печников А.А. Свидетельство о гос. регистрации программы для ЭВМ «Программа сбора вебметрических индикаторов по данным Google» Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ № 2014613369 от 25 марта 2014 г.
7. Печников А.А. Об измерениях вебметрических индикаторов // Международный журнал экспериментального образования. – 2013. – № 10. Часть 2. – С. 400-405.
8. Aguillo I.F., Granadino B., Ortega J. L., Prieto J. A. Scientific research activity and communication measured with cybermetric indicators // Journal of the American Society for the Information Science and Technology. 2006. 57(10). P. 1296 - 1302.
9. Огийко А.А. Применение метода Борда в ранжировании веб-сайтов // Процессы управления и устойчивость. 2014. Т.1(17). С. 357–362.
10. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа. М.: Финансы и статистика, 1983. 304 с.

# ОНТОЛОГИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ КОЛИЧЕСТВЕННОЙ СПЕКТРОСКОПИИ

А.И. Привезенцев, А.Ю. Ахлестин, Н.А. Лаврентьев,  
Н.Н. Лаврентьева, А.З. Фазлиев

*Институт оптики атмосферы СО РАН, Томск*

## 1. Введение

Параметры спектральных линий применяются в разных предметных областях: оптике атмосферы, атмосферной радиации, астрономии и т.д. Потребность в таких данных постоянно растет, и увеличивается число производителей экспертных данных. Более жесткими становятся требования к качеству данных: их точности, полноте и согласованности, достоверности и доверию. Показано [1], что для ряда прикладных задач существующие экспертные данные не адекватны требованиям, предъявляемым к ним, т.к. содержат в себе устаревшие, сомнительные и не полные наборы данных.

В ИОА СО РАН совместно рядом европейских университетов [2] создана информационная система (ИС) W@DIS [3]. Поскольку число молекул и их изотопологов, необходимых для решения прикладных задач, уже давно больше тысячи, а число опубликованных статей с параметрами спектральных линий составляет несколько десятков тысяч, назначением этой системы является формирование полной коллекции опубликованных спектральных данных и ее компьютерного описания.

Результаты анализа данных коллекции представляются в форме OWL-онтологий и каждая из них предназначена для ответа на соответствующую группу вопросов. К первой группе относятся вопросы о качестве источников данных, извлеченных из публикации. Среди вопросов этой группы такие вопросы как удовлетворяет ли конкретный источник данных правилам отбора, или насколько хорошо этот источник данных коррелирует с другими источниками данных, которые содержат переходы идентичные с частью тех, которые размещены в исходном источнике данных. Часть вопросов относится к экспертным источникам данных и связана с оценкой доверия к ним. Вторая группа вопросов содержит вопросы о качестве конкретных состояний или переходов молекулы. Третья группа вопросов связана с вопросами качества входных данных, используемых для решения задач спектроскопии.

При контроле качества данных используются ограничения, обусловленные математическими моделями молекул, и ограничения на допустимые значения корреляций. Они формируют основу для проверки достоверности. На практике традиционно используются экспертные данные, которые, как правило, являются компиляцией расчетных данных и данных измерений. Методы создания таких данных не являются аналитическими и по этой причине необходима оценка доверия экспертным данным.

## 2. Предметная область «Количественная спектроскопия»

При анализе предметной области (ПО) базовым вопросом является вопрос о том, какие предметные области входят в композицию соответствующую анализируемой ПО. В нашей работе в композицию входят три предметные области: «Вещество», «Спектроскопия» и «Информационные ресурсы». С какой степенью детализации описываются входящие в композицию предметные области? На какие вопросы можно получить ответы в онтологиях такой композиции предметных областей?

Разнообразие структур данных в спектроскопии большое. В количественной спектроскопии оно меньше и определяется главным образом квантовыми числами, определяющими уникальность состояний и переходов. Структура квантовых чисел зависит от симметрии молекулы, числа атомов, входящих в молекулу и выбора базисного набора функций при решении уравнения Шредингера. Математические модели молекул определяют ряд ограничений на состояния и правила отбора для переходов.

Стоит отметить, что последние 5 лет каждый год только в абсорбционной спектроскопии вычислены миллиарды переходов и сотни тысяч состояний для малоатомных молекул. По нашей оценке число измеренных переходов в год составляет десятки тысяч. Число журналов, в которых публикуется информация о спектральных данных, более сотни. Число статей по спектроскопии ежегодно публикуемых в журналах, трудах конференций, в сети интернет и т. д. - несколько тысяч (из них несколько сотен по спектроскопии высокого разрешения).

Изменчивость спектральных данных связана как с уточнением значений физических величин, так и с изменением числа физических характеристик, входящих в структуры данных. Дополнительные физические характеристики появляются при использовании новых моделей контура спектральной линий или использования новых наборов квантовых чисел. Сложной проблемой является переход от одного набора квантовых чисел к другому.

### 2.1. Вещество

Описательных моделей предметной области «Вещество» достаточно много. Не останавливаясь на их обзоре, мы рассмотрим только часть этой предметной области, связанную с моделью данных. В предлагаемой модели формализованы только изолированные атомы и молекулы и оставлены без описания вещества в разных фазовых состояниях. Эта часть может содержать свойства атомов и молекул, которые не относятся к спектроскопическим свойствам, но их значения являются входными данными для решения задач спектроскопии. В первую очередь к этим данным относятся данные о модельной структуре вещества.

## 2.2. Спектроскопия

Предметная область «Спектроскопия» ориентирована на описание спектральных свойств молекул, в первую очередь тех свойств, которые характерны для процессов испускания и поглощения излучения в атмосферах планет. Отличительная особенность нашей модели [4] состоит в том, что она основана на выделении первичных опубликованных источников данных, содержащих решение прямых и обратных задач, которые характеризуют значения физических величин, относящиеся к процессам испускания и поглощения. Число задач, решаемых в молекулярной спектроскопии, достаточно велико. Их можно разбивать на группы. Поскольку мы ограничились свойствами, относящимися к процессам испускания и поглощения, нас интересует группа задач, которые связаны с измерениями или вычислениями параметров спектральных линий, необходимых для описания этих процессов. Эти задачи образуют структуру, состоящую из двух цепей [5].

## 2.3. Информационные ресурсы

В этой предметной области нас интересует тот этап работы исследователя при котором информация о предметной области поступает к исследователям из опубликованных статей, написанных на естественном языке. Для исследователя технической задачей при работе со статьей является построение субъектно-предикатных структур [6], с целью получения выводов о результате исследований. В таких структурах элементарным информационным объектом является высказывание. Более сложными информационными объектами являются «источник данных» и «источник информации». Источник данных содержит в себе решение одной из задач спектроскопии и представляет собой часть публикации. Источник информации содержит в себе свойства решения относящегося к соответствующему источнику данных.

Перечень свойств решения задачи определяется исследователем исходя из информационных задач, которые ему необходимо решать. В нашей работе таких задач две. Это задача семантического поиска и задача автоматического построения экспертного массива данных. Заметим, что первичные источники информации, относящиеся к одной публикации, не содержат идентичных высказываний. Различие между публикацией и первичным источником информации может быть существенно меньшим по сравнению с различием между публикацией и первичным источником данных. Различие обусловлено теми свойствами решения задачи в публикации, которые вошли в определение того или иного источника информации. Например, такими свойствами могут быть описание достоверности решения задачи или описание оценки доверия экспертному массиву данных. Более того, высказывания, содержащиеся в первичном источнике информации, могут не содержаться в публикации.

Базовой проблемой информационных ресурсов, используемых в ИС, является контроль их качества. Для контроля качества используются критерии

достоверности данных, оценки доверия к данным и метрики, характеризующие свойства данных и классификаторов.

## **2.4. Качество экспертных данных в спектроскопии**

Экспертные информационные ресурсы количественной спектроскопии должны обеспечивать низкий порог трудозатрат для адекватного восприятия знаний о спектрах в прикладных науках. Последнее является важным в силу того, что исследователи прикладных наук не обладают знаниями необходимой глубины для понимания всех сторон спектроскопических данных, и их выбор строится чаще на доверии, а не на собственной проверке достоверности данных. По этой причине исследователи должны иметь полную картину о том, каким критериям достоверности и доверия удовлетворяют экспертные данные, полный набор результатов проверок по этим критериям. В нашей работе анализ качества экспертных спектральных данных сосредоточен на проверке их достоверности и оценке доверия по критерию опубликования [7, 8]. Такой анализ является частью решения задачи оценки доверия экспертных информационных ресурсов по критерию опубликования.

## **3. Онтологическое представление количественной спектроскопии**

В ИС W@DIS используется формальное описание предметных областей выполненное с помощью языка разметки OWL 2 DL. Детали построения онтологии информационных ресурсов по спектроскопии и онтологии состояний и переходов даны в наших работах [9-11]. Ключевыми конструктами языка OWL являются индивиды, свойства и классы.

Три предметные области, описанные выше, иницируют три группы вопросов, обусловленные качеством данных (см. Введение). С каждой из этих групп связаны определенные структуры индивидов (см. [9,10]). Первая группа вопросов, связанная с качеством публикаций по спектроскопии, порождает четыре типа индивидов: три типа связаны с задачами спектроскопии, а четвертый тип - с оценкой качества публикаций содержащих описание экспертных спектральных данных. Вторая группа вопросов относится к качеству каждого перехода или состояния, которые многократно измерены. Третья группа вопросов характеризует качество входных данных использованных при решении задач спектроскопии. Каждой группе вопросов соответствует свои онтологии. Стоит заметить, что для каждой молекулы можно строить независимые от других молекул онтологии. В подавляющем большинстве информационных задач спектроскопистов этими связями можно пренебречь. В таком приближении модульность онтологий достаточно высокая [8]. На Рис.1 показана модульность онтологии информационных ресурсов для решений задач спектроскопии. Здесь под термином «модульность» понимается возможность разбиения онтологии на множество независимых частей, каждая из которых может представляться отдельным файлом.

Выделение групп вопросов и построение связанных с ними онтологий является первым этапом формирования описания качества данных в коллекции ИС W@DIS. Следующим этапом является достижение онтологического соглашения с целью построения онтологии количественной спектроскопии [12].

Качество создаваемых онтологий для каждой молекулы существенно зависит от того насколько полна коллекция опубликованных данных и как хорошо согласованы данные разных источников. Заметим, что если для какой-то молекулы данные не полны и не согласованы, то оценка доверия экспертным данным для такой молекулы в рамках данной коллекции не должна проводиться.

В ИС W@DIS созданы приложения для автоматической генерации фактологической части онтологий при импорте источников данных в ИС.

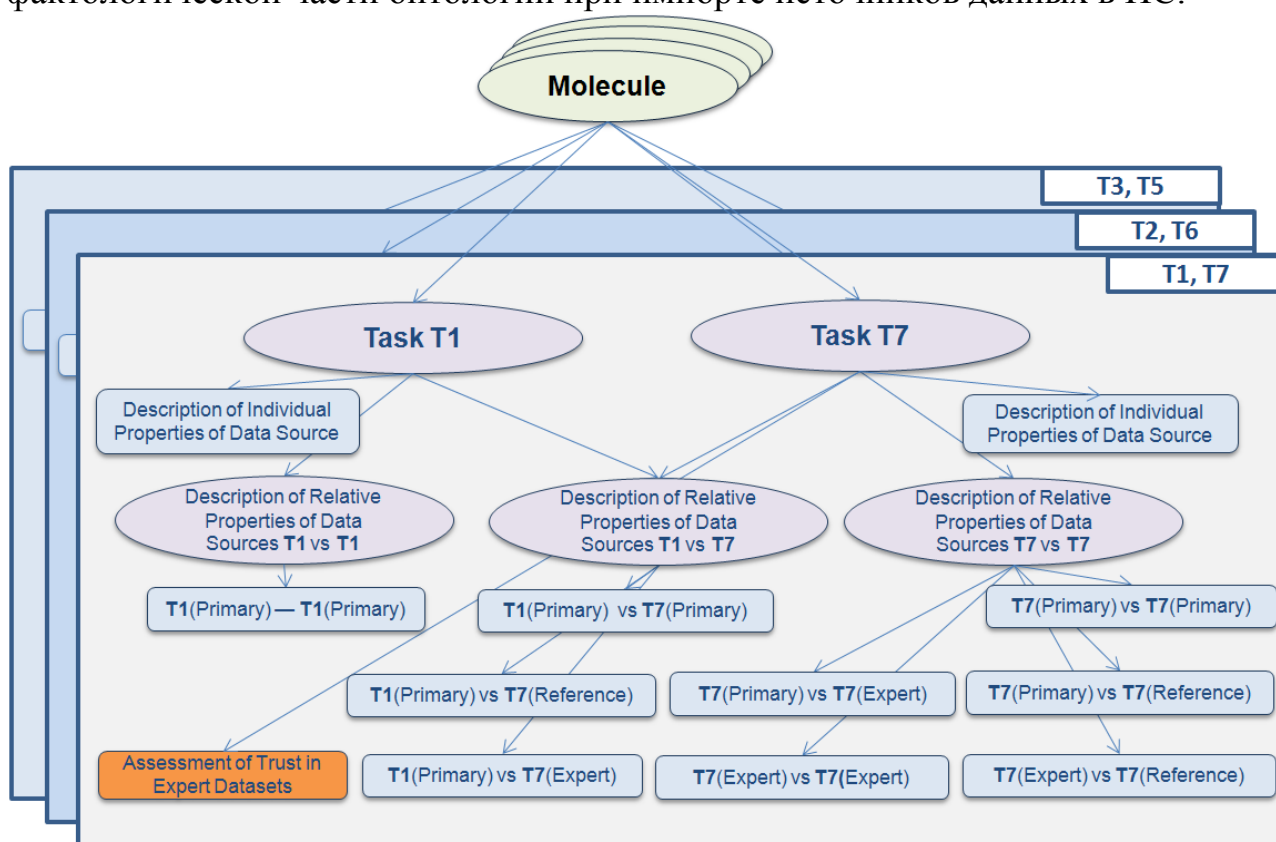


Рис. 1. Модульность онтологии на примере систематизации решений задач спектроскопии T1 и T7.

#### 4. Качество онтологий по спектроскопии

Контроль над качеством онтологий, формируемых автоматически возможен при условии измерения ее различных метрик. Метрики позволяют проводить быструю оценку онтологии и следить за ее изменениями. В нашей работе используются структурные метрики. Детальное описание метрик представлено в работе [11]. На Рис.2 показана таблица, содержащая число индивидов и количество высказываний для ряда изотопологов воды,

относящихся к онтологии информационных ресурсов по спектроскопии, на примере молекулы воды. Подавляющее число высказываний связано с описанием качества данных извлеченных из публикаций.

T-Box (Total: 3560 triple count, 166 class count, 72 object property count, 291 data property count)																				Catalog																							
Common individual count: 1293																				Ontology																							
Substance																																											
Substance	T1. Computation of states						T2. Computation of transitions						T3. Computation of collisional parameters						T5. Measurement of collisional parameters			T6. Measurement of transitions			T7. Computation of states																		
	VQN			Normal Modes			Schwenke			VQN			Normal Modes			VQN			Normal Modes			VQN			Normal Modes			Schwenke															
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C							
H <sub>2</sub> O	8	8	2	8	4	2	2	2	8	3	3	7	10	7	0	6	6	20	26	10	0	-	-	-	316	316	267	0	-	-	140	140	59	0	1	1	31	32	29	0	-	-	32
H <sup>18</sup> OH	3	3	2	4	4	1	1	1	4	4	4	4	8	7	0	4	3	15	19	18	0	0	-	-	61	61	53	0	-	-	77	77	43	0	-	-	18	18	18	0	-	-	18
H <sup>17</sup> OH	3	3	1	3	3	1	2	2	2	4	4	3	7	6	0	4	3	4	8	6	0	0	-	-	33	33	32	0	-	-	55	55	36	0	-	-	18	18	15	0	-	-	18
HOD	1	1	1	2	1	0	1	1	1	2	2	5	7	6	0	3	3	6	9	7	0	0	-	-	18	18	17	0	-	-	107	107	55	0	-	-	32	32	28	0	-	-	32
H <sup>18</sup> OD	-	-	1	1	1	0	-	-	1	2	2	2	4	3	0	2	2	3	5	4	0	0	-	-	17	17	17	0	-	-	27	27	26	0	-	-	9	9	8	0	-	-	9
D <sub>2</sub> O	-	-	2	2	2	0	-	-	2	3	3	1	4	3	0	3	3	2	5	4	0	0	-	-	9	9	8	0	-	-	41	41	30	0	-	-	19	19	10	0	-	-	19
D <sub>2</sub> <sup>18</sup> O	-	-	1	1	1	0	-	-	1	2	2	1	3	2	0	2	2	2	4	3	0	0	-	-	3	3	3	0	-	-	6	6	5	0	-	-	5	5	5	0	-	-	5
H <sup>17</sup> OD										2	2	1	3	3	0	2	2	1	3	3	0	0	-	-	14	14	14	0	-	-	13	13	11	0	-	-	3	3	3	0	-	-	3
D <sub>2</sub> <sup>17</sup> O										2	2	0	2	2	0	2	2	1	3	3	0	0	-	-	2	2	2	0	-	-	2	2	2	0	-	-	1	1	1	0	-	-	1
HTO										-	-	1	1	0	0	-	-	1	1	0	0										-	-	2	2	0	0							
Total	15	15	10	21	16	4	6	6	19	24	24	25	49	39	0	28	26	55	83	58	0	0	0	0	473	473	413	0	0	0	470	470	267	0	1	1	136	137	117	0	0	0	137

Statements/ Problems A-Box [Individual] Count	T1. Computation of states	T7. Computation of states	T2. Computation of transitions	T6. Measurement of transitions	T3. Computation of collisional parameters	T5. Measurement of collisional parameters
H <sub>2</sub> O	Source 3776 [807]	2269 [462]	7828 [3207]	19609 [5643]	14807 [6125]	57930 [29028]
	Decomposition		1	21997 [3726]	1	33957 [6340]
	Binary Relations	58627 [10294]		281929 [50833]		1452072 [243366]
H <sup>18</sup> OH	Source 1678 [457]	853 [133]	30535 [13307]	5205 [769]	32839 [14149]	21168 [15995]
	Decomposition			14359 [2437]		12580 [2454]
	Binary Relations	8018 [1653]		108380 [20282]		420317 [71343]

Рис.2. Статистика источников данных по молекуле воды и связанных с ними числа высказываний в онтологиях информационных ресурсов по молекуле воды.

## 6. Заключение

В работе рассмотрены два онтологические представления предметной области «Количественная спектроскопия». Первое из них характеризует качество публикаций из которых извлечены спектральные данные, а второе – качество измеренных характеристик переходов и состояний. Заметим, что шумовые данные, неопределенности, связанные с квантовыми числами и размерностями физических величин, неполнота информации о квантовых числах фиксировались на этапе формирования онтологии (графа знаний), но для ряда молекул нам не удалось в настоящее время достигнуть желаемого качества спектральных данных. Созданные онтологии описывают качество данных содержащихся в коллекции и ориентированы на работу с агентами.

Дальнейшие перспективы развития онтологии количественной спектроскопии мы связываем с развитием онтологий, характеризующих модели веществ, спектральные свойства которых входят в коллекцию ИС W@DIS.

Работа поддержана грантом РФФИ № 13-07-00411.



## ЛИТЕРАТУРА

1. Lavrentyev N.A., Makogon M.M., Fazliev A. Z., Comparison of the HITRAN and GEISA Spectral Databases Taking into Account the Restriction on Publication of Spectral Data // *Atmos. Ocean. Optics*, 2011. — V. 24. No. 5. PP. 436–451.
2. G. Császár, J. Tennyson, A.Fazliev, W@DIS - Prototype information system for systematization of spectral data of water // *Abst. 12-th Colloq. on High Resolution Molecular Spectroscopy (Dijon, 2007)*. —Dijon.: 2007. — P. 270-271.
3. W@DIS (Water Internet @ccesible Distributed Information System), URL: <http://wadis.saga.iao.ru>
4. Лаврентьев Н.А., Привезенцев А.И. Фазлиев А.З., Базы знаний для описания информационных ресурсов в молекулярной спектроскопии 2. Модель данных в количественной спектроскопии // *Электронные библиотеки*. 2011. — Т. 14. В.2.
5. Быков А.Д., Науменко О.В., Сеница Л.Н., Родимова О.Б., Творогов С.Д., Тонков М.В., Фазлиев А.З., Филиппов Н.Н., Информационные аспекты молекулярной спектроскопии — Из-во ИОА СО РАН, Томск. 2008. — 356С.
6. Зиновьев А.А. Основы логической теории научных знаний. — М.: Наука, 1967. — 260С.
7. Ахлестин А.Ю., Лаврентьев Н.А., Привезенцев А.И., Фазлиев А.З. Базы знаний для описания информационных ресурсов в молекулярной спектроскопии. 5. Качество экспертных данных // *Электронные библиотеки*, 2013. — Т. 16. В.4.
8. Fazliev, A. Privezentsev, D. Tsarkov, J. Tennyson, Ontology-Based Content Trust Support of Expert Information Resources in Quantitative Spectroscopy // In book: *Knowledge Engineering and the Semantic Web, Comm. in Comp. and Inform. Sci.*, V. 394, Springer Berlin Heidelberg, 2012.— PP.15-28.
9. Привезенцев А.И., Царьков Д.В., Фазлиев А.З. Базы знаний для описания информационных ресурсов в молекулярной спектроскопии 3. Базовая и прикладная онтологии // *Электронные библиотеки*, 2012. — Т. 15, В.2.
10. Voronina S.S., Privezentsev A.I., Tsarkov D.V., Fazliev A.Z., An Ontological Description of States and Transitions in Quantitative Spectroscopy // *Proc. of SPIE XX-th Intern. Symposium on Atmos. and Ocean. Optics: Atmos. Physics*, 2014. — V. 9292, 92920С.
11. Воронина С.С., Привезенцев А.И., Царьков Д.В., Фазлиев А.З., Различие онтологических представлений предметной области // *Труды конф. XVI Всеросс. научная конф. "Электронные библиотеки: перспективные методы и технологии"*, Дубна, ОИЯИ, 2014. — С. 124-130.
12. Guarino N. Formal Ontology and Information Systems // *Proceedings of FOIS'98*. Trento. Italy, 6–8 June 1998. Amsterdam: IOS Press. — P. 3–15.
13. Tsarkov D., Horrocks I., FaCT++ Description Logic Reasoner: System Description // *3rd International Joint conference on Automated Reasoning (IJCAR-2006)* — 2006.

# ИНФРАСТРУКТУРА ВЕБ-СЕРВИСА ИНТЕЛЛЕКТУАЛЬНОЙ ПРОБЛЕМНО-ОРИЕНТИРОВАННОЙ СИСТЕМЫ НАУЧНОЙ АНАЛИТИКИ ПО ФИЗИЧЕСКОЙ ХИМИИ РАДИКАЛЬНЫХ ЖИДКОФАЗНЫХ РЕАКЦИЙ В ИНТЕРНЕТ

А.И. Прохоров<sup>1</sup>, Д.А. Варламов<sup>1,2</sup>, Е.С. Амосова<sup>1</sup>,  
М.Е. Соловьева<sup>1</sup>, В.Е. Туманов<sup>1</sup>, П.К. Берзигияров<sup>1</sup>

<sup>1</sup> *Институт проблем химической физики РАН, г. Черноголовка, РФ*

<sup>2</sup> *Институт экспериментальной минералогии РАН, г. Черноголовка, РФ*

Разработка проблемно-ориентированных веб-сервисов в рамках создания интеллектуальных информационных систем по прогнозированию физико-химических свойств соединений методами интеллектуального анализа экспериментальных кинетических и термодинамических данных на основе онтологических моделей является актуальной междисциплинарной задачей.

Решению задачи извлечения опубликованных данных о химических реакциях и константах скорости реакций на основе онтологических моделей в биологии посвящен ряд работ зарубежных исследователей [1-3]. Методология и общие подходы к разработке инфраструктуры проблемно-ориентированных веб-сервисов, направленных на интеллектуальный анализ больших объемов информации, обсуждались в работах [5-8].

Целью настоящей работы является описание прототипа программно-технологической инфраструктуры проблемно-ориентированного веб-сервиса для автоматического сбора и семантического анализа кинетических и термодинамических данных из открытых оцифрованных источников в Интернет с последующей их разбраковкой, сохранением и использованием в интеллектуальной проблемно-ориентированной системе научной аналитики по физической химии радикальных реакций, которая является развитием проблемно-ориентированной системы по физической химии радикальных реакций [9] за счет внедрения в нее семантики на основе онтологических моделей предметной области.

В настоящее время информационная система [9] имеет многослойную программно-технологическую архитектуру как показано на рис. 1.

Верхний уровень системы представляет собой предметно-ориентированное веб-приложение, которое предоставляет пользователю интерфейс доступа к программным компонентам системы: информационной подсистеме, аналитической подсистеме, подсистеме дистанционного обучения, встроенной подсистеме объяснений и подсистеме производства новых профессиональных знаний.

Ниже доступного для пользователей уровня скрыт уровень программных агентов, распределенных на различных узлах локальной сети. В системе

реализованы интеллектуальные агенты, реактивные агенты, агенты обученных искусственных нейронных сетей, которые реализуют работу встроенных в портал экспертных систем и выполняют функции поиска информации. Набор интеллектуальных агентов системы является активным компонентом системы. Их можно представить в виде специализированных сервисов. При этом сами агенты ориентированы на обработку научных данных в узкоспециализированном разделе предметной области.

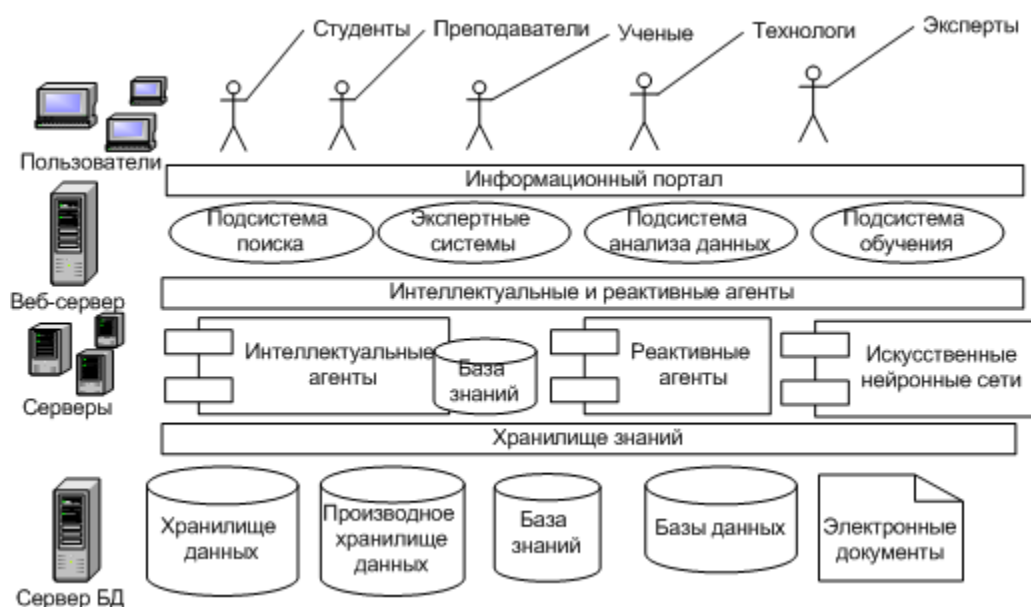


Рис. 1. Программно-технологическая архитектура системы научной осведомленности.

Вся система опирается на уровень хранилищ системы. Фактологическая часть хранилищ включает в себя:

- хранилище данных по константам скоростей радикальных жидкофазных реакций (накоплено более 30000 значений);
- киоск данных по энергиям диссоциации связей молекул (более 1000 объектов);
- база данных по энтальпиям образования радикалов (около 980 объектов)
- база данных по энтальпиям образования органических молекул (экспериментальные данные, до 3000 информационных объектов).

Хранилище знаний настоящей системы содержит эмпирические и рассчитанные факты, продукционные правила и процедуры расчета, а в совокупности с экспертными системами образует виртуальную подсистему производства новых профессиональных знаний. Дополнительно в качестве механизмов производства новых знаний в системе используются обученные искусственные нейронные сети (для реакций  $R\bullet + RH$ ,  $C_6H_5\bullet + RH$  и  $H + RH$  в водных растворах). Например, для хранилища по константам скорости в системе реализованы две экспертные системы (ЭС). Одна – классическая на

продукционных правилах (многоагентная), другая – на основе искусственных нейронных сетей. Решение об использовании конкретной ЭС определяет специализированный интеллектуальный агент. Если решение задачи пользователя можно определить расчетным путем на основе уже имеющихся данных, то управление по расчету нужной величины передается соответствующему резистивному агенту.

Дальнейшее развитие системы мы видим в увеличении интеллектуальной составляющей следующим образом – преобразование уровня хранилища системы в уровень хранилища знаний посредством реализации для системы ряда онтологических моделей (онтологий).

В [10] разработана одна из онтологических моделей представления радикальных жидкофазных реакций для внедрения ее в создаваемую систему (рис. 2).

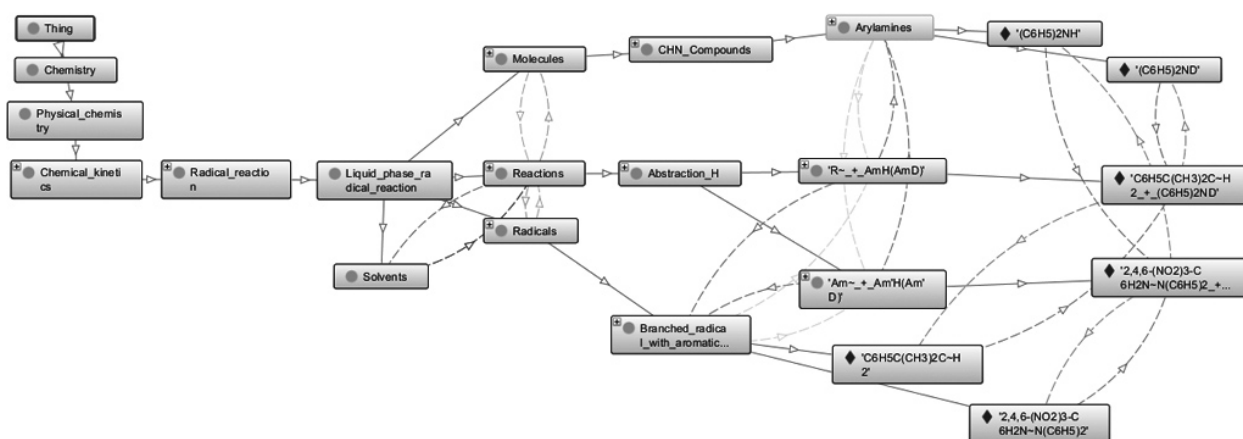


Рис. 2. Графическое представление фрагмента онтологии предметной области «радикальные реакции в жидкой фазе», отображающее классы и подклассы, а также связи между классами и экземплярами классов.

Добавляя в систему онтологические модели, с их помощью мы связываем все имеющиеся данные хранилищ, витрин, баз на уровне их измерений в единое хранилище знаний. При этом имеющиеся в системе продукционные правила необходимо согласовать с правилами, получаемыми из онтологий. Поэтому часть «старых» продукционных правил может перейти в разряд аксиом или фактов.

К этому моменту наша система практически будет готова к семантическому поиску в ней. Но в данный момент подобный сервис в системе отсутствует, поэтому предполагается его реализовать, поставив ему еще и дополнительную задачу: искать не только внутри системы, но еще и во вне неё. Добывая дополнительно данные извне, мы расширяем количество фактического материала и его актуальность, и, как следствие, достоверность предсказаний.



Рис. 3. Программно-технологическая архитектура веб-сервиса для автоматического сбора кинетических и термохимических данных радикальных жидкофазных реакций из открытых оцифрованных источников в Интернете.

Предполагается, что основными задачами прототипа веб-сервиса для автоматизации сбора кинетических и термохимических данных из открытых оцифрованных источников в Интернете будут:

1. поиск информации (поисковый профиль «ключевые слова», список сайтов (журналов), в перспективе семантический поиск);
2. формирование блока документов (форматы pdf, doc, htm, xml), либо автономно (из offline-коллекций), либо через Интернет;
3. парсинг документов (извлечение данных из текстовых документов: семантическая модель документа, кластеризация, классификатор (байесовы сети), искусственные нейронные сети, нечеткие нейросети);
4. верификация и сохранение полученных в хранилище данных системы.

Таким образом, действующая информационная система [9] будет обновлена до интеллектуальной проблемно-ориентированной системы научной аналитики по физической химии радикальных жидкофазных реакций. Её «опорная» часть «Хранилище знаний» пополнится дополнительными фактологическими составляющими: онтологическими моделями по кинетике и термохимии радикальных жидкофазных реакций и, связанным с ними, хранилищем оцифрованных документов, которые предполагается собирать, в том числе и из открытых источников в Интернет. Кроме того, всё хранилище знаний системы организуется вокруг разрабатываемых онтологических моделей в единое целое. Это и наличие в онтологиях предметных понятий и связей между ними позволяет реализовать в системе подсистему семантического поиска как в материалах хранилища знаний системы, так и в Интернет. Семантический поиск и сбор документов в Интернет будет осуществляться автоматизированной частью веб-сервиса с последующим

интеллектуальным анализом и обработкой полученных документов. Основной задачей системы останется возможность предсказания значений кинетических и термодинамических характеристик радикальных жидкофазных реакций.

Проведение работ по данной тематике поддержано грантом РФФИ № 15-07-08645-а, рук. Прохоров А.И.

## ЛИТЕРАТУРА

1. Weber L., Irmer M., Bobach C. Extraction of chemical reactions from full text documents: From n-tuples of value attribute pairs toward the automated construction of reaction databases [Электронный ресурс] / [http://www.chemaxon.com/wp-content/uploads/2014/07/B05\\_Weber\\_OntoChem\\_ACS\\_2014.pdf](http://www.chemaxon.com/wp-content/uploads/2014/07/B05_Weber_OntoChem_ACS_2014.pdf) (дата обращения 20.05.2015 г.)
2. Hakenberg J., Schmeier S., Kowald A., Klipp E., Leser U. // *OMICS A Journal of Integrative Biology*, 2004. Vol. 8, 2. P.131-152.
3. Jessop D.M., Adams S.E., Murray-Rust P. Mining chemical information from open patents // *Journal of Cheminformatics*. Vol. 3. № 40. P. 1-17.
4. Bingjun Sun, Qingzhao Tan, Prasenjit Mitra, C. Lee Giles Extraction and Search of Chemical Formulae in Text Documents on the Web / International World Wide Web Conference Committee (IW3C2). WWW 2007, May 8.12, 2007, Banff, Alberta, Canada. [Электронный ресурс] / <http://www.sciweavers.org/read/extraction-and-search-of-chemical-formulae-in-text-documents-on-the-web-37295> (дата обращения 20.05.2015 г.)
5. Wild D.J. A Web Service Infrastructure for Chemoinformatics // 4th Joint Sheffield Conference on Chemoinformatics, June 2007. [Электронный ресурс] / <http://cisrg.shef.ac.uk/shef2007/talks/wild.pdf> (дата обращения 20.05.2015 г.)
6. Wild D.J. Integrating text and literature sources with traditional chemoinformatics tools. [Электронный ресурс] / <http://acscinf.org/docs/meetings/233nm/presentations/233nm07.pdf> (дата обращения 20.05.2015 г.)
7. Viroli M., Casadei M., Nardini E., Omicini A. Towards a Pervasive Infrastructure for Chemical-Inspired Self-organising Services // *Self-Organizing Architectures. Lecture Notes in Computer Science*. 2010. Vol. 6090. P. 152-176.
8. Ai C. S., Blower P.E., Jr., Ledwith R. H. Extraction of chemical reaction information from primary journal text. // *J. Chem. Inf. Comput. Sci.* 1990. Vol. 30, № 2. P. 163–169.
9. Туманов В.Е., Прохоров А.И., Лазарев Д.Ю., Соловьева М.Е. Система научной осведомленности по физической химии радикальных реакций. // *Информационные ресурсы России*. 2010. № 5. С. 16-21.
10. Амосова Е.С., Туманов В.Е. Представление химических реакций, реагентов и их термодинамических свойств в интеллектуальной системе по физической химии радикальных реакций в жидкой фазе с использованием онтологической модели предметной области // *Бутлеровские сообщения*. 2014. Т. 39. № 7. С. 39-46.

# СИСТЕМАТИЗАЦИЯ ФИЗИКО-ХИМИЧЕСКИХ ДАННЫХ. ВОЗМОЖНОСТИ ОНТОЛОГИЙ И БАЗ ДАННЫХ, ПРИНЦИП ДОПОЛНИТЕЛЬНОСТИ

В.А. Серебряков<sup>1</sup>, А.О. Еркимбаев<sup>2</sup>, В.Ю. Зицерман<sup>2</sup>, Г.А. Кобзев<sup>2</sup>,  
К.Б. Теймуразов<sup>1</sup>, Р.И. Хайрулин<sup>2</sup>

<sup>1</sup> *Вычислительный центр им. А.А.Дородницына РАН*  
<sup>2</sup> *ОИВТ РАН*

Накопление и систематизация физико-химических данных и прежде всего данных о свойствах вещества является одним из основных направлений деятельности в естественнонаучных дисциплинах. До недавнего времени основным информационным ресурсом этой области была традиционная БД. Однако резкий рост их числа при многообразии форматов и моделей привел к тому, что действующая инфраструктура оказалась неспособной обеспечить эффективную организацию рабочего процесса. Автономность БД с жесткой фиксацией используемых терминов и логических структур практически исключает возможность беспрепятственного обмена данными без активного участия человека-эксперта. В итоге интерес в научном сообществе сместился в сторону принятия онтологий как более эффективных средств формализации и распространения научных данных. Главное назначение онтологий в том, что они добавляют к данным семантику (смысл, правильное толкование) и соотношения, что в совокупности описывает «знание» с возможностью его машинной интерпретации. Оценить богатый потенциал, заложенный в онтологиях для хранения и распространения данных, можно, сопоставляя их возможности с БД [1-3]. При кажущемся сходстве решаемых задач между ними имеются глубокие различия. Онтология служит для распространения информации, определяя на формальном языке концепции и соотношения, которые представляют содержание и структуру предметной области. В то же время концептуальная схема БД, определяя все понятия и структуру данных, служит только для тех целей, что реализует конкретная БД. Распространение информации посредством онтологии проводится согласованным образом, то есть передаваемая ею структура данных является общедоступной и одинаково трактуемой в определенном сообществе. Все члены сообщества могут использовать онтологию и имеют доступ к информации.

Авторы ряда публикаций специально выделили все факторы, определяющие сходства и различия обеих конструкций. Их сводка, зафиксированная в лекции [2], приведена в таблице 1.

Таблица 1.

Концептуальная схема БД	Онтология
Определяет структуру БД на формальном языке	Определяет набор концепций и соотношений, которые представляют содержание и структуру предметной области на формальном языке
Фокусируется на данных	Фокусируется на смысле
Сущности	Классы
Атрибуты	Соотношения
Ограничения	Аксиомы
Нет таксономии	Таксономия – ключевой элемент онтологии
Данные – ключевой элемент	Экземпляры данных не обязательны
Семантика только в концептуальной схеме, спроектирована для человека, не эволюционирует с изменением БД и приложений	Семантика – ключевой элемент, доступный программной обработке
Схему трудно изменять и поддерживать	Потенциально легче изменять и поддерживать

Наряду с семантикой, которая теряется в ходе проектирования БД, в таблице указаны некоторые дополнительные признаки. В частности, БД немислима без содержания в виде данных, в то время, как в онтологии наличие экземпляров не обязательно (*instances optional*). С другой стороны, онтология немислима без таксономии классов (хотя и не сводится к ней), а в БД таксономия как структурный компонент отсутствует. В итоге, формальный характер онтологии позволяет реализовать машинные выводы и рассуждения, что совершенно не предусмотрено в концептуальной схеме БД. Применительно к тематике «свойства веществ и материалов» особую роль играет возможность на уровне онтологий поддерживать эволюцию схемы данных, связанную с расширением круга объектов и появлением новых, ранее неизвестных понятий (последний пункт в табл.). Так, эволюция схемы может быть связана с постепенным включением наноструктур в химические БД [4, 5].

Поскольку БД отличается высокая производительность при поиске и реализации сложных запросов, недостижимая для других архитектур, в качестве основной стратегии выбрано не «вытеснение» БД, а создание своеобразной связки путем так называемого *database-to-ontology mapping*. Ее задача — использовать преимущества обеих конструкций, в основном за счет соединения семантики с высокой производительностью при работе с данными. В известном смысле, выбранный путь следует «принципу дополнительности»



при использовании обеих концепций, обладающих в чем-то взаимоисключающими свойствами.

Авторы [3], исходя из того, какая из двух концепций (онтология или схема БД) может рассматриваться как ведущая, выделили две стратегии. Первая из них ориентирована на использование онтологий для усиления функциональности БД: реализации запросов с использованием согласованной в сообществе специалистов семантики; облегчения проектирования БД; обеспечения интеграции нескольких БД. Вторая стратегия, напротив, рассматривает БД как возможный инструмент усиления или даже разработки онтологии.

Соответственно можно выделить две концепции взаимного отображения БД и онтологии. Первой соответствует БД, основанная на онтологии (**DBBO**, *database based on ontology*), когда смысл сущностей, записанных в БД, определен ссылкой на соответствующую онтологию. Вторая концепция предполагает проектирование структуры, получившей название онтологии, основанной на БД (**OBDB**, *Ontologies Based on DB*). Задача такой структуры — разместить экземпляры онтологии в БД, чтобы обеспечить достаточную эффективность при загрузке, поиске и реализации сложных запросов. В основном эта концепция нацелена на решение задач Semantic Web, где требуется организовать хранение и управление архивами документов, записанных на языке **RDF**, в то время как первая концепция (**DBBO**) обращена, в основном, к сообществу специалистов по БД.

Ниже речь будет идти о системах типа **DBBO**, включающих: исходно принятую БД с заполняющими ее записями; онтологию для семантического индексирования БД; возможные ссылки на другие онтологии с целью расширения словаря; соотношения между каждым элементом БД и онтологическим понятием. Онтология в такой структуре лишь обеспечивает семантику, но не содержит экземпляров — их роль выполняют записи БД.

Усиление функциональности за счет онтологии возможно не только для существующей БД, но и на этапе проектирования, что заметно ускоряет разработку и снимает ряд возникающих ограничений. В частности, концептуальные схемы, предлагаемые разными экспертами, в известной степени унифицируются, поскольку базируются на единой модели области. Кроме того, уже на этапе проектирования снимается проблема семантики, поскольку каждый элемент данных получается отображением онтологических понятий, имеющих точное определение и смысл.

Наряду с проектированием БД, использование онтологий открывает возможности в интеграции БД, ранее созданных без ссылок на онтологии. Онтология предоставляет достаточно эффективный способ объединения схем, позволяя преодолеть проблемы с отсутствием семантики в исходных БД. Пользователь может формулировать в терминах онтологии запросы, которые будут обращены к «подключенным» к онтологии БД. Тем самым онтология будет играть роль эффективного посредника между пользователем и данными.

Один из наиболее впечатляющих примеров расширения функциональности БД за счет связывания с онтологией дает химическая БД низкомолекулярных соединений содержания **ChEBI** (Chemical Entities of Biological Interest Ontology) [5], которая охватывает молекулярные сущности, их группы и классы. Первое из понятий относится к любым идентифицируемым по составу атомам, молекулам, наноструктурам и т.п., второе — группу связанных атомов (или один атом) в составе молекулы (methyl, CH<sub>3</sub>), третье — (класс) совокупность молекул или групп, охваченных классификационным признаком, например alkane (R-CH<sub>3</sub>). Каждую сущность, а также группу и класс, идентифицирует уникальный идентификатор **ChEBI ID** (например, **ChEBI:15377** для воды), свободный для цитирования в сети пользователем или программным агентом.

В записи для молекулярной сущности (рис. 1) указаны химические данные (определение, формула, масса, заряд), названия, как принятое в **ChEBI**, так и синонимы из других источников, структурная информация, кодированная в линейных нотациях (**InChI** и **SMILES**), а также детализированная в **Molfile** (координаты атомов и матрица связности). Наряду с этим в записи представлены регистрационные номера в химических классификаторах (CAS, Weilstein и др.) и ссылки на другие БД с обширной информацией, например термодинамической в БД NIST ([webbook.nist.gov/chemistry/](http://webbook.nist.gov/chemistry/)) или токсикологической в БД ChemIDplus ([chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp](http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp)).

The screenshot displays the ChEBI entry for paracetamol (CHEBI:46195). Key elements include:

- Recommended ChEBI name:** paracetamol
- ChEBI ID:** CHEBI:46195
- Definition:** A derivative of phenol which has an acetamido substituent located *para* to the phenolic -OH group.
- Chemical structure:** A benzene ring with a hydroxyl group (-OH) at the bottom and an acetamido group (-NH-C(=O)-CH<sub>3</sub>) at the top.
- Chemical structure searches:** Options to find compounds containing or resembling this structure.
- Additional chemical data:**
  - InChI:** 1S/C8H9NO2/c1-6(10)9-7-2-4-8(11)5-3-7/h2-5,11H,1H3,(H,9,10)
  - InChIKey:** RZVAJINKPMORJF-UHFFFAOYSA-N
  - SMILES:** CC(=O)Nc1ccc(O)cc1
  - Formula:** C<sub>8</sub>H<sub>9</sub>NO<sub>2</sub>
  - Net Charge:** 0
  - Mass:** 151.16260
- Links to the ChEBI ontology:** A section showing relationships like 'paracetamol (CHEBI:46195) has role cyclooxygenase 2 inhibitor (CHEBI:50629)'.

Рис. 1. Внешний вид типовой записи в БД **ChEBI**.

Главным элементом онтологии, интегрированной с БД, является таксономия, причем любое понятие может происходить от нескольких родительских понятий. Внизу рис. 2 показаны дочерние (*incoming*) и родительские (*outgoing*) классы, непосредственно связанные с веществом в записи (в данном случае **paracetamol, C<sub>8</sub>H<sub>9</sub>NO<sub>2</sub>**). Обращение к URI с указанием онтологии и уникального ID непосредственно связывает внешний ресурс (или пользователя) с соответствующей записью, например URI [www.ebi.ac.uk/chebi/searchId.do?chebiId=46195](http://www.ebi.ac.uk/chebi/searchId.do?chebiId=46195) связывает его с записью в БД для парацетамола.

Отдельные классы связаны соотношениями, либо типа класс-субкласс (таксономия), либо ассоциативными соотношениями, которые определяют свойства или роли отдельных понятий, см. табл. 2.

Таблица 2.

Таксономические соотношения	<b>Is_a</b>
	<b>Has_part</b>
Ассоциативные соотношения	<b>Is_conjugate_base_of</b>
	<b>Is_conjugate_acid_of</b>
	<b>Is_tautomer_of</b>
	<b>Is_enantiomer_of</b>
	<b>Has_functional_parent</b>
	<b>Has_parent_hydride</b>
	<b>Is_substituent_group_of</b>
	<b>Has_role</b>

Первые два (**is\_a**, **has\_part**) определяют простейшие виды логической связи, восемь остальных — фиксируют химический контекст в отношениях двух веществ, например, соотношение **has\_functional\_parent** (как видно из рис. 2) указывает, чем является *paracetamol* по отношению к другому веществу *paracetamol sulfate*. Особое значение в списке имеет последнее соотношение **has\_role**, позволяя раскрыть множество аспектов в свойствах и приложении веществ. Оно связывает базовую суб-онтологию **Molecular structure**, в рамках которой построена таксономия веществ, с другой суб-онтологией **Role**, которая включает три класса верхнего уровня (**biological role**, **chemical role**, **application**). При этом **chemical role** классифицирует сущности по их «химически значимой» роли (кислота, основание, лиганд), **biological role** определяет их «биологически значимую» роль, а **application** дает классификацию по целевому использованию, например **drug**, **pesticide** и др. Все сущности, относящиеся к этой суб-онтологии, так же как и вещества, имеют уникальный ID, что позволяет связывать внешний ресурс с любой из сущностей в онтологии **role**. Пользователь может совершать навигацию по

дереву, соответствующему ролевой онтологии, точно так же, как и для веществ, включенных в БД.

Таким образом, наличие онтологии открывает для пользователя или внешнего ресурса возможности, исходно отсутствующие в БД. В частности, по идентификатору **ChEBI** можно сделать ссылки на любые сущности БД, определить их логические и ролевые связи, встроенные в многоуровневые таксономии. В результате можно возложить на программного агента реализацию сложных запросов с поиском веществ по структуре, функциям, биологической роли и т.п. В целом, это означает, что наряду с данными для конкретного вещества, **ChEBI** предоставляет фрагмент «знаний», характеризующих предметную область.

Общая концепция **DBVO** была опробована ранее при создании БД по теплофизическим свойствам веществ [6]. Роль онтологии здесь особенно существенна, поскольку наряду с унификацией семантики необходимо поддерживать эволюцию схемы данных и аксиомы, отражающие логические и математические ограничения, присущие данной области. Ключевое понятие — набор данных, включающий для одного вещества несколько констант и температурных функций, а также сведения о фазовом состоянии вещества, единицах измерений, неопределенности и источнике данных. Основные списки — веществ, свойств, фазовых состояний, единиц измерений и т.д. считаются открытыми, что позволяет в рамках онтологии поддерживать эволюцию схемы данных.

Концептуализация предметной области привела к выбору 12 базовых понятий, послуживших основой для построения соответствующих классов. Среди них группа базовых классов (вещества, состояния, свойства, численные данные), 6 вспомогательных классов (размерности, неопределенность, источник данных и др.) и 2 класса, определяющих вычисляемые функции и аргументы. При этом класс **Functions** порождает 2 суб-класса, определяющих значения свойств и выполнение математических ограничений, определяемых требованиями предметной области, например равенство энергии Гиббса для сосуществующих фаз (жидкость-газ или жидкость-твердое тело).

При разработке онтологии активно привлекались внешние источники (существующие онтологии и словари) для унификации семантики. В частности, для именованых веществ использован словарь **ChemSpider**, который обеспечивает присвоение веществу уникального идентификатора, например для водорода **CSID:762** и соответствующего URI [www.chemspider.com/Chemical-Structure.762.html](http://www.chemspider.com/Chemical-Structure.762.html). Отдельные термины, связанные с фазами, свойствами, размерностями приняты из онтологий **ChemAxiom** и **QUDT** (*Quantities, Units, Dimensions and Data Types*) [4].

Физические принципы, определяющие свойства веществ, накладывают целую совокупность ограничений на использование понятий. Логические ограничения записаны с использованием конструкций языка OWL. Среди логических ограничений — разбиение класса свойств на два непересекающихся

класса (свойства-функции и свойства-константы); обязательность определения аргумента для свойства-функции; согласованность ссылок на состояния вещества с видом свойств-функций (например, запрет на свойство **viscosity** в состоянии **solid**). Математические ограничения задаются отдельно для каждого экземпляра-свойства. Они относятся к свойствам-функциям: требования к области определения, области существования, характеру монотонности и возможно другим характеристикам (например, связи двух и более функций). Математические ограничения касаются не классов, а экземпляров свойств, поскольку области определения и существования функций определяются для каждого свойства отдельно и список свойств допускает расширение. На основании построенной онтологии сгенерирована реляционная БД в СУБД PostgreSQL. Соответственно 12 классам онтологии создано 12 Java классов, которые отображаются на таблицы реляционной БД данных, рис. 2.

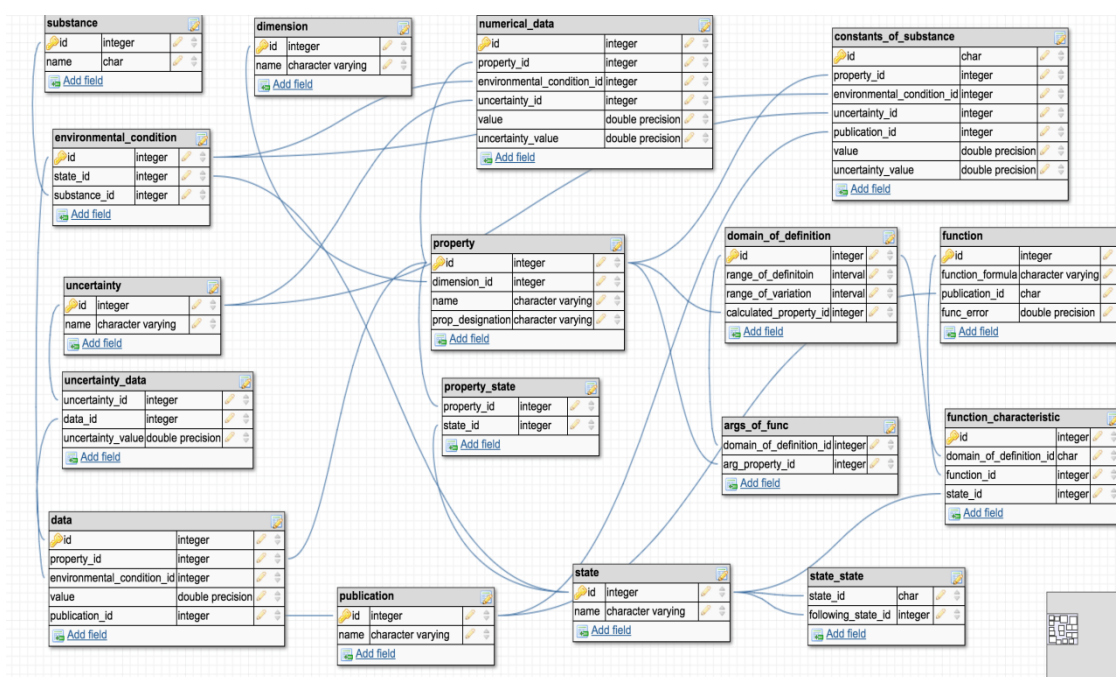


Рис. 2. Таблицы теплофизической БД [6].

Разработанная система заметно облегчила проектирование реляционной БД за счет связи с онтологией. Помимо унификации семантики, онтология обеспечила выполнение логических и математических связей между понятиями и сохранение за пользователем права наращивать списки веществ, свойств, единиц измерения и прочих элементов набора данных при эволюции концептуальной схемы.

В работе изучен один из новых подходов к работе с данными по свойствам, использующий технологии связывания БД с онтологиями. Сами по себе БД, будучи созданы в разных коллективах, с неизбежностью порождают разноречивой терминологии, логических схемах и форматах данных. Онтология оказалась идеальным средством унификации семантики и подчинения

множества концептуальных схем единой структуре, к тому же согласованной в научном сообществе. В результате открывается путь к широкой интеграции структурно или даже тематически разнородных ресурсов с возможностью обмена данными и использования в совместной работе.

Работа выполнена при поддержке РФФИ – проект № 13-07-00218.

## ЛИТЕРАТУРА

1. Когаловский М.Р., Калиниченко Л.А. Концептуальное и онтологическое моделирование в информационных системах // Программирование. — 2009. — Т. 35. — №5. — С. 3-25.
2. Uschold M. Ontologies and Database Schema: What's the Difference? 2011. URL: [www.slideshare.net/UscholdM/](http://www.slideshare.net/UscholdM/)
3. Laallam F.Z., Kherfi M.L., Benslimane S.M. A survey on the complementarity between database and ontologies: principles and research areas// Int. J. Computer Applications in Technology. — 2014. — V. 49. — No 2. — P. 166–187.
4. Зицерман В.Ю., Кобзев Г.А., Фокин Л.Р. Возможности и перспективы информационных технологий в подготовке и распространении справочных данных: свойства веществ и материалов // Научно-техническая информация. Серия 1. — 2004. — №2. — С. 7-14.
5. Degtyarenko K., de Matos P., Ennis M. et al. **ChEBI**: a database and ontology for chemical entities of biological interest// Nucleic Acids Research. — 2008. — V. 36. — Database issue. — D344–D350.
6. Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Практическая реализация системы интеграции теплофизических данных на основе онтологической модели предметной области // ИНФРАСТРУКТУРА НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ И СИСТЕМ: Труды Четвертого Всероссийского симпозиума (С.-Петербург. 6-8 октября 2014 г.). — М.: ВЦ РАН, 2014. — С. 415-421.

# ИССЛЕДОВАНИЕ "ЭЛЕКТРОННОГО УЧАСТИЯ" С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ И ТЕХНОЛОГИЙ АВТОМАТИЗИРОВАННОГО МОНИТОРИНГА ВЕБ-РЕСУРСОВ

А.В. Чугунов<sup>1</sup>

*1 Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО)*

**Аннотация.** Для эффективного решения задач внедрения технологий электронного правительства и методов электронного управления, в том числе электронного участия, необходимы регулярные исследования не только готовности общества к внедрению новых технологий, но и мониторинг получаемых эффектов от реализации проектов. В статье представляются промежуточные результаты работы, осуществляемой в Центре технологий электронного правительства Университета ИТМО с целью создания комплекса программных компонентов, обеспечивающих специализированный сервис для исследователей, изучающих процессы общественного участия и интерактивного взаимодействия власти с гражданами и бизнесом в Российской Федерации. Представлен пример использования автоматизированной системы мониторинга портала "Российская общественная инициатива" для проведения научного исследования.

**1. Введение: актуальность исследования.** В Российской Федерации с 2006 года регулярно предпринимаются усилия по расширению использования информационно-коммуникационных технологий (ИКТ) для реинжиниринга деятельности органов власти (административная реформа), оказания государственных услуг и обеспечения интерактивного взаимодействия с гражданами. Мировой опыт показывает, что эффективное внедрение технологий электронного управления (E-Governance) может предоставлять гражданам и бизнесу доступ к высококачественным услугам госорганов, уменьшить стоимость и базовую трудоемкость оказания этих услуг, а также организовать механизм «обратной связи» между властью и обществом.

В 2009 - 2010 гг. в России в качестве основной задачи в рамках создания электронного правительства были обозначены планы перевода государственных и муниципальных услуг в электронный вид. В это же время происходило формирование и уточнение приоритетов государственной программы «Информационное общество (2011 - 2020 гг.)». К сожалению, в качестве основных критериев успешности выполнения планов перевода услуг в электронный вид стали рассматриваться формальные показатели ввода в действие электронных сервисов и подключения к системе электронного межведомственного взаимодействия, а не улучшения качества оказания услуг.

Эти показатели нашли отражение и в рейтингах, которые в настоящее время являются одним из инструментов государственной политики в сфере развития электронного правительства. Ряд аспектов, связанных с рейтингованием и оценкой качества реализации проектов электронного правительства были проанализированы в научных публикациях [1; 2; 3], аналитических обзорах и докладах на научных конференциях [4], в том числе международных [5; 6]. Важно что критический анализ проблем, связанных с информатизацией госсектора, осуществляют не только исследователи из академической среды, но и правительственные аналитические центры [7].

Тем самым можно констатировать, что, несмотря на то, что электронные услуги и сервисы создаются в веб-среде, практически отсутствуют адекватные механизмы автоматического мониторинга эффективности этого процесса. Представляется, что отсутствие обратной связи и соответствующих административных действий в результате выявления неэффективности функционирования государственных информационных ресурсов и систем, является одним из факторов низкой отдачи от вкладываемых государством ресурсов в развитие электронного правительства.

Однако в настоящее время довольно активно развивается новый для России инструмент электронного управления (E-Governance) – электронное общественное участие (E-Participation). Этот сегмент общественной жизни (а также соответствующий инструментарий) формируется силами не только и не столько государственных структур, а общественными организациями и движениями, а также самими гражданами. На политическом уровне технологии электронного участия нашли поддержку в программах и проектах под названием «Открытое правительство». В случае, когда речь идет о развитии технологий электронного участия, появляется возможность анализа данных, получаемых непосредственно с информационных ресурсов, оказывающих соответствующие сервисы. Информационная система, представляемая в данной статье, предназначена для аккумуляции данных порталов электронного участия, их обработки и представления как сервис для исследователей.

**2. Электронное участие как междисциплинарное исследовательское направление и практика в Российской Федерации.** В мировой практике развитие механизмов электронного участия граждан в жизни государства (E-Participation), принятие политических решений – уже далеко не новая тема для исследований. В частности, анализ мирового потока научных публикаций по базе Scopus показывает устойчивый рост количества статей в международных научных журналах и индексируемых трудах конференций, начиная с 2005 года.

В начале 2013 г. в России тема «электронное участие» (иногда под названием «электронная демократия», что не совсем корректно с научной точки зрения) стала предметом дискуссий не только на государственном уровне, но и в научно-образовательном и экспертном сообществе. Анализ научных исследований и публикаций позволяет резюмировать, что развитие механизмов



электронного участия – глобальная тенденция. При этом для каждой страны существуют своя история и динамика появления таких инструментов, во многом обусловленные социальной, институциональной и технологической готовностью всех участников процесса взаимодействия [8].

На сегодняшний день в России функционирует довольно много общественных площадок для размещения электронных петиций и голосования. Наиболее значимые из них:

- Change.org (<http://www.change.org/ru>);
- OnlinePetition (<http://www.onlinepetition.ru>);
- Демократор.ру (<http://democrator.ru>);
- Наше мнение (<https://mypetition.ru>);
- Просто россияне (<http://prosto-rossiane.ru>);
- Демократия 2 (<http://democratia2.ru>);
- Народная инициатива КПРФ (<http://ni.kprf.ru>).

В марте 2013 года Указом Президента Российской Федерации В.В. Путина утверждены Правила рассмотрения общественных инициатив, направленных гражданами Российской Федерации с использованием интернет-ресурса «Российская общественная инициатива» [9]. Данным Указом деятельность по работе с инициативами граждан была распространена на региональный и муниципальный уровни. Этим же Указом была определена некоммерческая организация (Фонд информационной демократии - <http://www.f-id.ru/>), уполномоченная на осуществление функций, предусмотренных Правилами рассмотрения инициатив. Сам факт назначения оператора портала электронных петиций Указом Президента является весьма знаковым свидетельством, что на тот момент данная тема имела высокий приоритет на уровне главы государства.

С самого старта портала электронных петиций «Российская общественная инициатива» (РОИ - <http://www.roi.ru>, т.е. с апреля 2013 г., Центр технологий электронного правительства Университета ИТМО осуществляет его текущий мониторинг. В течение первого года мониторинг проводился без использования автоматизированных методов, т.к. на портале РОИ отсутствовали соответствующие интерфейсы. Ситуация изменилась в апреле 2014 г. после открытия на портале API (Application Programming Interface). Тем самым для сбора данных с портала РОИ начал использоваться официальный API, который позволяет получить данные обо всех инициативах и их текущих состояниях (название, описание, количество голосов, статус, прогресс, даты начала и окончания голосования и др.) [10].

Портал РОИ имеет четкое разделение адресности инициатив: федеральный, региональный или муниципальный уровни. С первых дней стала понятна базовая тенденция – основная масса (около 90%) всех инициатив ориентирована исключительно на федеральный уровень. Муниципальные и региональные инициативы составляют 5% и 6% соответственно от общего количества размещенных инициатив. Имеется возможность отслеживать

тематическое распределение петиций и другие аспекты функционирования портала РОИ.

**3. Автоматизированная система мониторинга порталов электронного участия как сервис для исследователей.** Первый компонент автоматизированной системы ориентирован на анализ и обработку данных, получаемых с портала петиций РОИ. Основой для постановки задачи и разработки явилась деятельность, которая осуществлялась аналитиками Центра технологий электронного правительства Университета ИТМО в течение года (апрель 2013 – апрель 2014) в «ручном режиме». Появление на портале РОИ интерфейса API позволило автоматизировать процесс сбора первичных данных.

Рассмотрим архитектурные и другие решения, лежащие в основе созданной системы мониторинга (рис. 1).

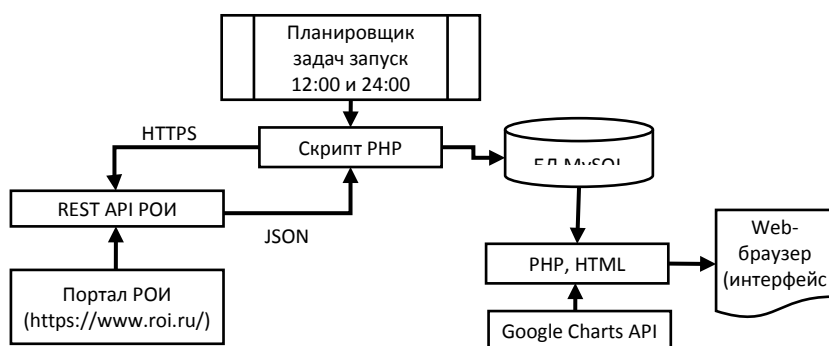


Рис. 1. Архитектура программного модуля «Мониторинг Российской общественной инициативы - РОИ»

Ниже представлена структура базы данных (используется MySQL) модуля «Мониторинг РОИ» (рис. 2).



Рис. 2. Структура базы данных MySQL модуля «Мониторинг Российской общественной инициативы - РОИ»

Система мониторинга (модуль мониторинга РОИ) в настоящее время обеспечивает следующие возможности для исследователей:

а) представление данных о динамике голосов (возможный минимальный временной интервал - час):

- общее количество голосов за день;
- количество голосов по каждой инициативе;

б) мониторинг текущего состояния (прогресс голосования во времени по положительным и отрицательным голосам, отслеживание изменения статуса по инициативам);

в) использование фильтров и срезов для поиска инициатив (со ссылкой на портал РОИ) и для анализа текущего состояния и динамики голосования:

- уровень инициативы;
- регион;
- статус;
- категории;
- даты начала и окончания голосования;
- голоса и соотношение положительных и отрицательных голосов (в т.ч. динамика роста, прогресс);
- название, описание – текстовый контент.

Использование возможностей автоматизированного мониторинга позволяет осуществлять проведение некоторых групп аналитических исследований:

- аналитика позитивного отклика (анализ контента по инициативам, которые набирают большое количество голосов «За» или больше положительных голосов, чем отрицательных) – создание облака тегов из названий инициатив и/или их содержания;

- аналитика отрицательного отклика (анализ контента по инициативам, которые набирают больше отрицательных голосов) – создание облака тегов;

- выявление проблемных вопросов в конкретных регионах на основе контентного анализа инициатив по регионам (в том числе инициатив федерального уровня);

- оценка интереса граждан к проекту РОИ (по количеству инициатив и сравнению динамики голосов).

Вся аналитика строится на основе базы данных (сведения в которую собираются через API 1.0 для автоматизированного взаимодействия с РОИ). Запросы к БД осуществляются на языке SQL, полученные данные (таблицы данных) можно использовать для дальнейшего анализа или оформления графических представлений как в редакторах электронных таблиц (Excel, OpenOffice Calc и другие), так и на web-ресурсах.

В настоящее время на основе опыта осуществления аналитических исследований с использованием автоматизированных средств, ведется проектирование программных модулей, призванных собирать данные и представлять их для исследователей по другим группам ресурсов, которые

относятся к категории «электронное участие». В первую очередь, это ресурсы, предоставляющие гражданам возможность направить жалобу или обращение в адрес органов власти или хозяйствующих субъектов по поводу каких-либо нарушений или действий, создающих проблемы в различных сферах. В первую очередь, это благоустройство, защита прав и свобод граждан, злоупотребления и прочие проблемы.

Команда исполнителей данного проекта планирует его развитие с целью обеспечения исследователей надежным сервисом на основе данных извлекаемых в реальном режиме времени с порталов «электронного участия».

**4. Пример использования системы для проведения прикладного политологического исследования.** Рассмотрим специфику использования механизмов «электронного участия» в политических целях и задачах политического PR. Естественно, что сразу после открытия портала петиций (РОИ), российская внесистемная оппозиция приступила к использованию его возможностей для давления на власть.

В частности группа А. Навального начала весьма активно размещать на портале РОИ свои петиции и использовать этот факт как информационный повод для консолидации своих сторонников, демонстрируя тем самым уровень общественной поддержки своих взглядов. В этой связи необходимо предостеречь от завышенных оценок, т.к., если подходить с точки зрения объективного политологического анализа, поддержка оппозиционной петиции, направленной на антикоррупционные меры (например, запрет чиновникам покупать автомобили элитного класса), отнюдь не означает, что гражданин, подписавший эту петицию, разделяет политические взгляды конкретной оппозиционной группы. Тем более, что если гражданин зашел на портал либо самостоятельно, либо через нейтральные ресурсы, то он даже и не узнает, что подписался под петицией, размещенной известным оппозиционером. Также гражданин, голосующий за конкретную инициативу, не имеет возможности увидеть, кто автор этой инициативы, соответственно без информационной поддержки и использования собственных каналов донесения информации о петиции до своих сторонников и сочувствующей аудитории, ее автор обречен на «анонимность».

Хорошим примером, демонстрирующим организованную PR-кампанию по сбору голосов в поддержку оппозиционной петиции, может служить весьма резонансный сюжет о петиции А. Навального «Об уголовной ответственности за незаконное обогащение чиновников и иных лиц, обязанных представлять сведения о своих доходах и расходах» (инициатива № 77Ф9376).

Инициатива была размещена на портале РОИ 9 января 2014 г. и ее продвижение заняло почти год. Весь процесс (рис. 3) можно разделить на 2 этапа: первый – эволюционный (январь-сентябрь 2014 г.), когда медленно и поступательно шел набор голосов; и второй этап – «пиаровский» (25 сентября –

декабрь 2014), который собственно и обеспечил для этой петиции прохождение 100 тысячного рубежа.

Динамика голосований за петицию во время первого этапа демонстрирует явно проигрышный вариант: за 5 месяцев было собрано всего около 3 тыс. голосов. Кампания PR-сопровождения сбора голосов стартовала 25 сентября 2014 г. после размещения на сайте А. Навального статьи, что привело к появлению 6 тыс. голосов в день публикации и на следующий день еще более 3 тыс. голосов. Для раскрутки был использован хештег #20 в твиттере, а преодоление 20-тысячной отметки в голосовании вызвало видимо существенное усиление активности сторонников данной инициативы. Далее PR-кампания стала нарастать, и в декабре были получены 100 тыс. голосов, что создало возможность вынесения инициативы на рассмотрение экспертной группы.

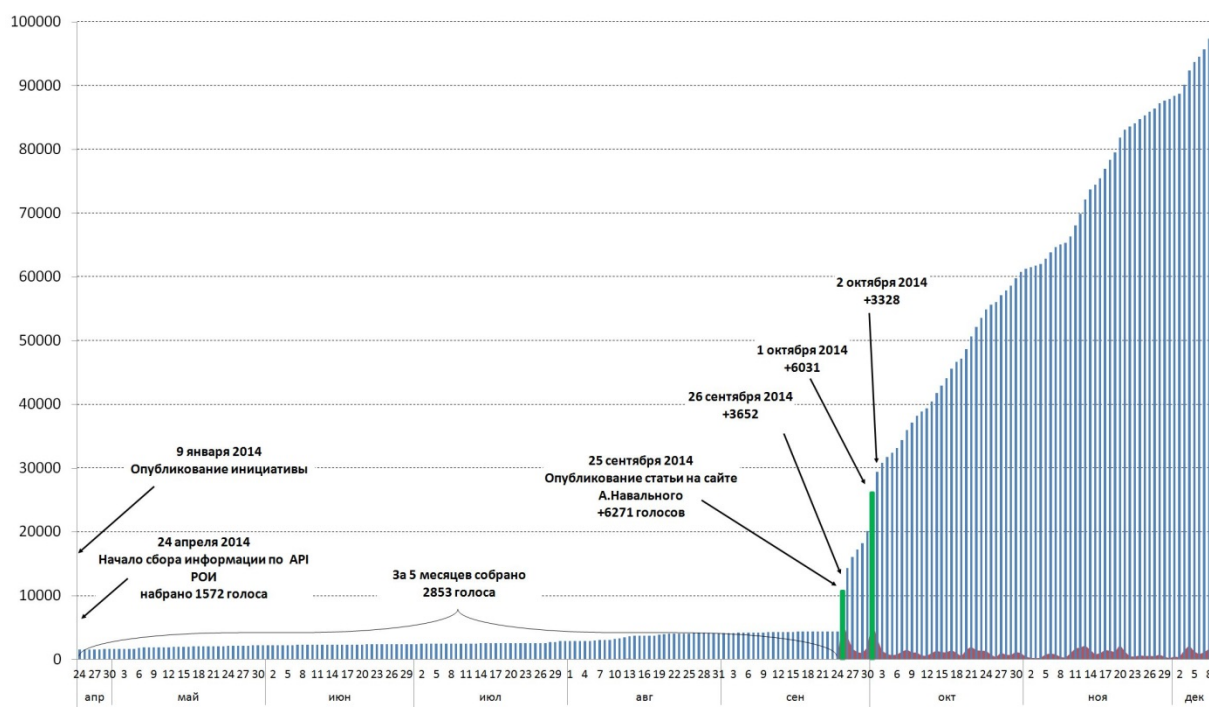


Рис. 3. Динамика голосования за инициативу «Об уголовной ответственности за незаконное обогащение чиновников и иных лиц, обязанных представлять сведения о своих доходах и расходах» (№ 77Ф9376) на портале РОИ, 2014 г. (по данным системы автоматизированного мониторинга Центра технологий электронного правительства Университета ИТМО)

Вопрос обсуждался 9 февраля 2015 г. на заседании экспертной рабочей группы федерального уровня, возглавляемой М. Абызовым. Петиция подавалась от Фонда борьбы с коррупцией (ФБК), возглавляемого А. Навальным. ФБК предлагал распространить юрисдикцию России на статью 20 Конвенции ООН против коррупции и ввести в российский Уголовный кодекс новый состав преступления «незаконное обогащение». Решение на

первом заседании не было принято в связи с отсутствием кворума. 26 февраля 2015 г. на следующем заседании экспертная рабочая группа решила не передавать в Госдуму инициативу против коррупции, признав при этом важной саму проблему уголовной ответственности чиновников за незаконное обогащение. Камнем преткновения для экспертов стало признание некоторых пунктов предлагаемого инициативой законопроекта противоречащими Конституции РФ. В качестве компромисса рабочая группа рекомендовала Парламенту и Правительству подготовить предложения по ужесточению ответственности за незаконное обогащение.

Следует отметить, что обсуждение инициативы А. Навального стимулировало выявление некоторых противоречий и нерешенных процедурных моментов в функционировании Российской общественной инициативы (регламентов принятия решений и прочих вопросов), что привело к необходимости подготовить специальный доклад в администрацию Президента по данной теме.

**5. Выводы и направления продолжения работ.** Результаты представленной работы позволяют сделать вывод о пригодности созданного программного комплекса к проведению многоаспектных исследований в социальной и политической сферах.

Расширение состава информационных ресурсов данного типа (порталы электронного участия), подключенных к системе мониторинга, требует наличия API, либо других интерфейсов, пригодны для организации прямого доступа к первичным данным. К сожалению разработчики сайтов не всегда предусматривают такую возможность.

Для эффективного решения задач внедрения технологий электронного правительства и методов электронного управления, в том числе «электронного участия», необходимы регулярные исследования не только готовности общества к внедрению новых технологий, но и мониторинг получаемых эффектов от реализации проектов. В статье представлены промежуточные результаты работы, осуществляемой в Центре технологий электронного правительства Университета ИТМО с целью создания комплекса программных компонентов, обеспечивающих специализированный сервис для исследователей, изучающих процессы «электронного участия» в Российской Федерации.

Работа выполнена при поддержке Российского гуманитарного научного фонда по проекту «Электронное взаимодействие между обществом и властью: исследование процессов социодинамики и институционализации» (проект № 13-03-00603).

## ЛИТЕРАТУРА

1. Чугунов А.В. Развитие Интернета в России и мониторинг общественного мнения о ходе реализации ФЦП «Электронная Россия» (2002 – 2006 гг.) // Информационные ресурсы России. 2007. № 4 (98). С. 21 – 27. URL: <http://elibrary.ru/item.asp?id=11592208>
2. Чугунов А.В. Мониторинг общественного мнения об использовании общественных пунктов доступа в Интернет и востребованности электронных правительственных услуг (2005 – 2007 гг.) // Информационные ресурсы России. 2008. № 1 (101). С. 2-7. URL: <http://elibrary.ru/item.asp?id=10021590>
3. Чугунов А.В. Развитие электронного правительства в России: международные рейтинги и актуализация текущих проблем реализации проектов на федеральном и региональном уровне // Межотраслевая информационная служба. 2013. № 2. С. 31-40. URL: <http://elibrary.ru/item.asp?id=19040428>
4. Бершадская Л.А., Чугунов А.В. Востребованность услуг электронного правительства: анализ дискуссий в социальных сетях // Интернет и современное общество: сборник научных статей. Труды XVI Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2013), Санкт-Петербург, 9 – 11 октября 2013 г. – СПб: НИУ ИТМО, 2013. С. 67 – 71. URL: <http://elibrary.ru/item.asp?id=21718254>
5. Bershadskaya L., Chugunov A., Trutnev D. Monitoring Methods of E-Governance Development Assessment: Comparative Analysis of International and Russian Experience // ACM International Conference Proceeding Series. ICEGOV 2012 - 6th International Conference on Theory and Practice of Electronic Governance: Proceedings. N.Y.: ACM, 2012. P. 490-491. URL: <http://elibrary.ru/item.asp?id=20478829>
6. Bershadskaya L., Chugunov A., Trutnev D. E-Government in Russia: Is or Seems? // ACM International Conference Proceeding Series. ICEGOV 2012 - 6th International Conference on Theory and Practice of Electronic Governance: Proceedings. N.Y.: ACM, 2012. P. 79-82. URL: <http://elibrary.ru/item.asp?id=20478851>
7. Хромов А.В. Информатизация госструктур: проблемы, способы решения, тенденции / Аналитический центр при Правительстве РФ. URL: [http://www.cnews.ru/reviews/ppt/forum\\_2013/gos/1\\_Hromov.pdf](http://www.cnews.ru/reviews/ppt/forum_2013/gos/1_Hromov.pdf)
8. Bershadskaya L., Chugunov A., Trutnev D. E-Participation Development: A Comparative Study of the Russian, USA and UK E-Petition Initiatives // ACM International Conference Proceeding Series. ICEGOV 2013 - 7th International Conference on Theory and Practice of Electronic Governance: Proceedings. Seoul; South Korea; 22-25 October. N.Y.: ACM, 2013. P. 73-76.
9. Указ Президента Российской Федерации от 4 марта 2013 г. № 183 «О рассмотрении общественных инициатив, направленных гражданами Российской Федерации с использованием интернет-ресурса "Российская

общественная инициатива"» (с изменениями и дополнениями). URL: <http://base.garant.ru/70326884/>

10. Чугунов А.В. Электронная демократия и развитие электронного участия в России: предварительные результаты мониторинга портала «Российская общественная инициатива» // Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты (ISC-14): Материалы Третьей Международной научно-практической конференции. 2014. С. 295-298. URL: <http://elibrary.ru/item.asp?id=22713218>



# ЛИЧНОСТЬ И ЭПОХА: ВИРТУАЛЬНОЕ ПРЕДСТАВЛЕНИЕ НАУЧНОГО НАСЛЕДИЯ

Д.М. Шакирова<sup>1</sup>, Д.Ш. Сулейманов<sup>1</sup>, О.А. Невзорова<sup>1</sup>, М.И. Курманбакиев<sup>1</sup>

<sup>1</sup> *Институт прикладной семиотики АН РТ*

## АННОТАЦИЯ

В условиях формирования глобального информационного общества чрезвычайно важным становится обеспечение доступа к социально-значимой информации, в первую очередь научного, образовательного и культурного характера. Сохранение научного наследия в современных информационных технологиях позволяет с новых позиций представить личность ученого через призму событий его эпохи, выделить наиболее значимые достижения и представить их в современном ракурсе восприятия. Такая задача ставится в проекте портала «Научные школы АН РТ», который будет содержать совокупность информационных ресурсов, представленных в виде информационных систем – виртуальных музеев-библиотек (ВМБ) ученых Академии наук РТ. Информационная система в виде виртуального музея-библиотеки будет включать архивные материалы - научные труды, статьи, переписки, творческие материалы, фото-видео материалы, воспоминания элиты нашего общества в разные исторические периоды - и современные информационные интерактивные ресурсы - форумы, видеоконференции, интерактивные модели, обучающие курсы.

## 1. ВВЕДЕНИЕ

Многообразие научных школ, языков, национальных культур и конфессий является уникальным историко-культурным наследием РТ и должно быть в полном объеме представлено в мировом информационном пространстве. В области науки, образования и культуры национальные достижения складываются из персональных достижений талантливых личностей, живших и творивших в определенную историческую эпоху и в конкретной социально-экономической и социально-культурной среде. Проект научно-образовательного портала «Научные школы Академии наук Республики Татарстан» ставит целью создать Интернет-ресурс, ориентированный на исследование объективных исторических, научных, культурологических характеристик общества через представление достижений и роли ученого в конкретный исторический период. В качестве модели реализации выбрана модель виртуального музея-библиотеки (ВМБ) ученого, в основе которой лежит максимально полная база знаний о личности ученого и его времени. Структура портала включает различные по составу, внутренней структуре и функциональным возможностям персональные виртуальные музеи -

библиотеки ученых АН РТ. Цели проекта носят общекультурный, образовательный и просветительский характер и призваны привлечь внимание научной общественности и особенно молодежной аудитории к научному потенциалу АН РТ.

Как отмечается в [1-3] виртуальные музеи - интегративное культурное явление, обладающее многими перспективными функциями, как правило, не свойственными традиционным музеям. Виртуальные музеи – это новая культурная форма интегративного характера, многофункциональный комплекс, существующий в сети Интернет и не являющийся результатом модернизации какой-либо прежней культурной формы, - виртуальные музеи несводимы к сайтам традиционных музеев. Феномен виртуальных музеев характеризуется тремя базовыми признаками: присутствием в виртуальном пространстве (наличие или отсутствие «физического» музея в данном случае не принципиально); наличием культурного продукта, репрезентирующего как модус прошлого, так и настоящего и будущего; адресацией широкому кругу лиц. Таким образом, главной целью проекта является создание нового (в том числе и по технологическим возможностям) культурного продукта, позволяющего отражать эпоху через события и достижения наиболее выдающихся личностей ученых.

## **2. КОНЦЕПЦИЯ ОТБОРА СОДЕРЖАНИЯ ВМБ**

Одной из самых сложных задач создания исторически достоверного, увлекательного, научно-обоснованного и полезного для людей разного возраста, типа, профессии виртуального музея является отбор информации. При построении концепции контента виртуального музея нами используются подходы, разработанные историками, библиографами, писателями для мемуарной литературы. Так, постсоветская мемуаристика, особенно ее «популярная» часть, впервые предложила, в отличие от обезличенной «правды» документа, личную историю, индивидуальный факт, биографию конкретных людей, живших в истории. Story на данном этапе главенствует над History, на смену универсальному опыту приходит опыт индивидуальный: история отдельной человеческой жизни сочетается с историей науки и общества. По-видимому, происходит некая приватизация истории, при которой мемуары, дневники, автобиографии, подборки научных книг, статей, фотографий, отзывов и воспоминаний современников предлагают читателю более достоверную версию прошлого, чем исторический документ.

С этих позиций особый интерес представляют появляющиеся в последние годы виртуальные музеи университетов с отдельными сайтами ученых, которые интегрируют музейную и библиотечную составляющие [4].

В настоящем исследовании разработана интегрированная контент - модель, которая в рамках историко-культурологического подхода может быть охарактеризована следующими основными признаками:

- авторская индивидуальность и субъективность текстов;

- сочетание воспоминаний с летописями;
- информационная насыщенность (использование материалов различных жанров: научные книги, публицистика, популярные статьи, публичные выступления);
- достоверность событий и характеристик упоминаемых известных личностей (наличие энциклопедических и биографических справок);
- визуализация событий и портретов людей на основе мультимедийных технологий;
- интерактивность (форумы, online конференции, образовательные модули);
- современность и перспективность (включение работ учеников, последователей научных идей, наложения актуальных событий).

### 3. АРХИТЕКТУРА И ОСНОВНЫЕ ФУНКЦИИ ВМБ

Архитектура научно-образовательного портала представлена на рис.1.

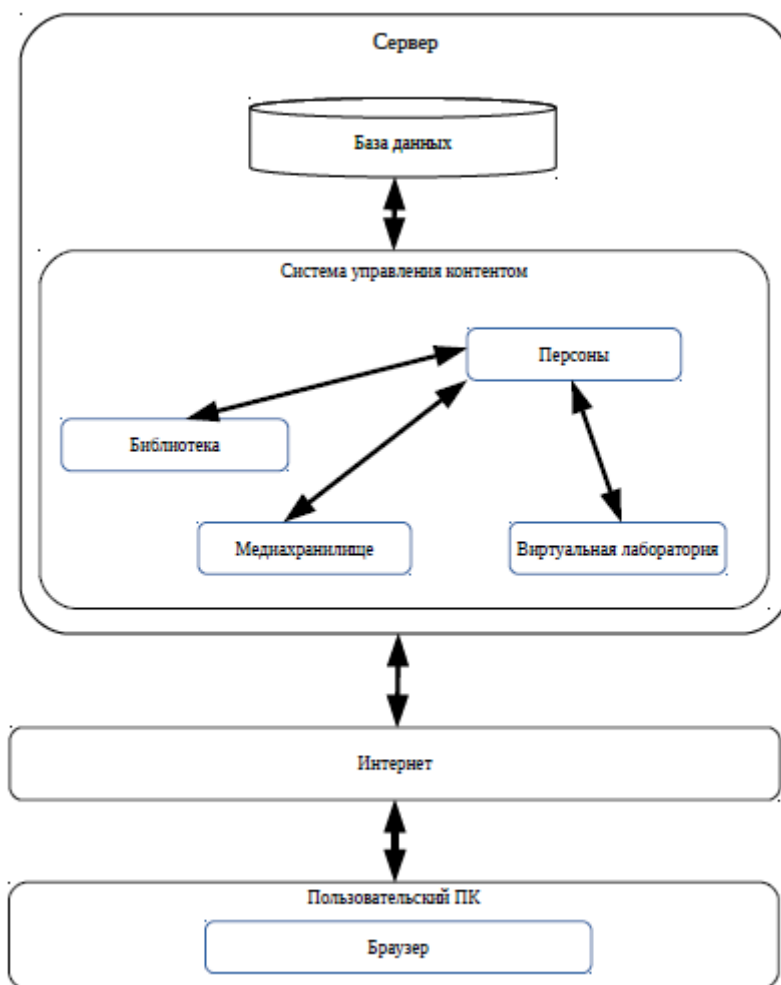


Рис.1. Архитектура научно-образовательного портала «Научные школы АН РТ»

На рис.1 представлены основные модули информационной системы, реализованной на основе архитектуры «клиент-сервер». Сервер приложений реализует прикладной компонент и обеспечивает компонент управления ресурсом. Компонент представления отвечает за пользовательский интерфейс. Система управления контентом (прикладной компонент) включает набор взаимосвязанных модулей: модуль «Персоны», модуль «Библиотека», модуль «Медиахранилище», модуль «Виртуальная лаборатория».

**Модуль «Персоны»** содержит информацию о персонах, представленных на портале.

Персональная информация структурирована по подразделам:

- Биографические сведения
- Научно-исследовательская работа
- Публикации
  - Статьи
  - Доклады
  - Книги
- Прочие сведения
- Цитируемость
  - В научных изданиях
  - В прессе
- Мультимедийная информация
  - Фотографии
  - Видео

**Модуль «Библиотека»** реализуется как электронный библиотечный каталог, а также хранилище полнотекстовых изданий, журнальных и газетных публикаций, мемуаров и личной переписки.

Библиотечные записи описываются следующими параметрами:

- Автор
- Название
- Тип
- Теги - цитируемость
- Теги - редакционная деятельность
- Доступ

Параметр «Тип» определяет тип публикации (книга, статья и др.).

В зависимости от значения параметра типа записи присваиваются различный набор параметров описания:

- |             |                  |
|-------------|------------------|
| Тип: книга: | Тип: статья:     |
| • ISBN      | • Журнал         |
| • ББК       | • ISSN журнала   |
| • УДК       | • номера страниц |
| • Тираж     |                  |

- Количество страниц
- Оглавление

Параметр «Доступ» определяет уровень доступа к публикации (только чтение; чтение-копирование; отсутствие доступа к содержимому).

Публикации, доступные для скачивания, предполагается представить в трех форматах: для чтения на компьютере и для чтения с различных мобильных устройств. Для чтения с ПК предполагается разработка функционала удобного онлайн-чтения непосредственно на сайте. Для обеспечения чтения с мобильных устройств материалы будут представлены в форматах FB2, EPUB. Форматы FB2 и EPUB являются специализированными форматами для электронных книг. Основное преимущество данных форматов перед различными текстовыми форматами заключается в возможности структурирования материала (по частям, главам и т. д.). Большинство программного обеспечения для чтения в этих форматах поддерживает формирование оглавления и навигацию по нему, а также поиск по ключевым словам и фразам, адаптацию внешнего вида документа под потребности пользователя (размер шрифта, цвет и т. д.), а некоторые программы поддерживают функцию голосового чтения.

Для аннотирования публикации используется ряд специальных тегов (тег-цитируемость, тег- редакционная деятельность). Первый тег относится к публикациям, которые цитируют автора, второй – связан с составом редколлегии издания.

**Модуль «Медиахранилище»** реализован как структурированный каталог медиа-файлов (фото и видео). Помимо организации стандартных альбомов планируется аннотировать медиа-файлы специальными семантическими тегами для реализации семантического поиска. В составе фотоколлекции содержатся редкие фотографии, некоторые из которых до начала работы над проектом существовали в единственном экземпляре в частных коллекциях.

**Модуль «Виртуальная лаборатория»** реализуется как научно-образовательная площадка с широкими функциями, такими как поддержка функционирования авторских интерактивных обучающих курсов; проведение семинаров и конференций; организация дискуссий и мастер-классов; презентация изданий и выставок и т.п.

Взаимосвязи данных, представленных в различных модулях, отражены на рис. 2. Такая схема взаимодействий позволяет автоматизировать процесс заполнения персональных кабинетов, а также наиболее удачно структурировать информацию в подразделах.

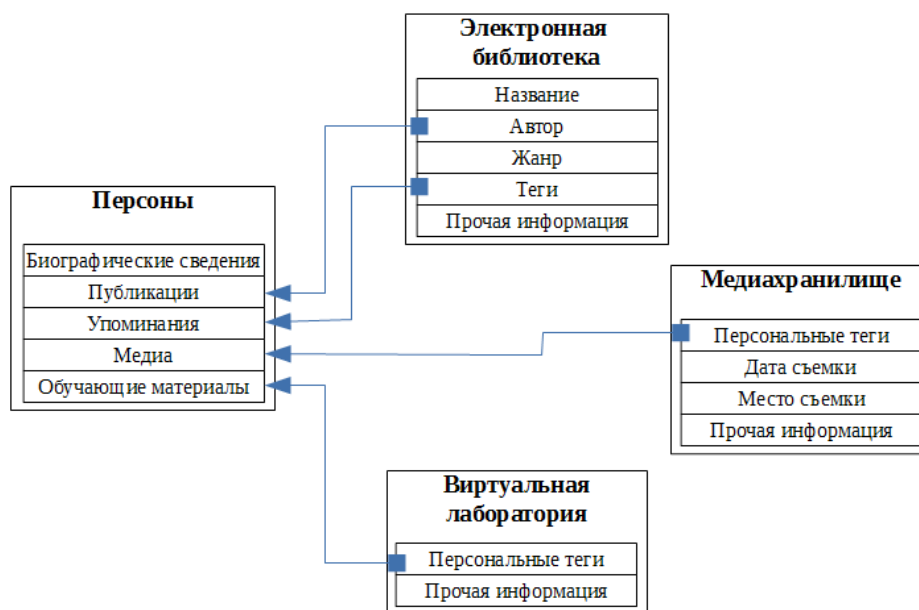


Рис.2. Схема взаимодействия данных различных модулей

Модель данных в разрабатываемой информационной системе позволяет реализовать многоаспектный поиск, в том числе семантический на основе семантических тегов. Перспективными направлениями является развитие основных поисковых функций, обеспечивающих многоязычный поиск по информационным единицам на основе многоязычных классификаторов, словарей и тезаурусов по различным научным направлениям; разработка лингвистического обеспечения для семантического аннотирования текстов на естественном языке; разработка новых интерактивных моделей общения (интерактивные обучающие среды, виртуальные научные лаборатории и др.).

#### 4. ВИРТУАЛЬНЫЙ МУЗЕЙ – БИБЛИОТЕКА АКАДЕМИКА М.И. МАХМУТОВА

В настоящей статье представлены основные результаты разработки виртуального музея-библиотеки крупного российского филолога-востоковеда, теоретика образования, общественного деятеля, академика АН РТ и РАО Мирзы Исмаиловича Махмута. Идея разработки концепции ВМБ, включающей биографические, научные и общественно-политические воззрения известного в России и за рубежом ученого энциклопедической направленности деятельности, академика М.И. Махмута в рамках исторической эпохи XX – начала XXI вв. представляется актуальной именно сейчас в период осмысления советского прошлого и российского настоящего. Интерес к педагогическим идеям М.И. Махмута не ослабевает и они активно используются в педагогической практике ряда стран (Канада, США, Китай, Сингапур и др.). Поэтому ВМБ является фактически виртуальной научной лабораторией и предназначена для активного развития педагогических идей Махмута его учениками и последователями во всем мире.

Благодаря наличию богатого видео ряда, фильмам и отрывкам реальных интервью ученого, высказываниям его коллег М. Махмутов раскрывается как реальная личность, и идеи выдающегося российского педагога, востоковеда, государственного деятеля становятся современными и живыми. ВМБ включает архивные материалы - научные труды, статьи, переписку, творческие материалы, фото-видео материалы, воспоминания элиты нашего общества в разные исторические периоды, а также современные информационные интерактивные ресурсы - форумы, видеоконференции, интерактивные модели, обучающие курсы.

#### **Состав научного наследия М.И. Махмутова:**

- каталог публикаций – **более 550 материалов;**
- электронная научная коллекция трудов и иного письменного наследия М.И.Махмутова на разных языках – **около 1000 источников;**
- тематические коллекции («М.Махмутов и современное образование», «М.Махмутов и национальное образование», «Последователи научной школы М.Махмутова в мире» и др.)

Электронная библиотека М.И. Махмутова включает электронные документы, сгруппированные по следующим разделам:

- биография (ключевые события, жизнь и деятельность ученого)
- воспоминания
- награды
- каталог книг (на разных языках)
- каталог публикаций (статьи, доклады, выступления)
- конференции
- пресса об ученом
- диссертации (кандидатские, докторские)
- интересы ученого
- галерея (мультимедийная коллекция - фото, видео и аудио архивы, в том числе и современных мероприятий).

Прототип информационной системы размещен на сервере Академии наук Республики Татарстан: [www.vml.antat.ru](http://www.vml.antat.ru). Вся информация на портале будет размещена на трех языках: русском, татарском и английском, что даст уникальную возможность представить личность Махмутова М.И. во всем мировом пространстве.

## **ЗАКЛЮЧЕНИЕ**

Научно-образовательный портал представляет собой системное многоуровневое объединение различных научных и образовательных ресурсов и сервисов. Цели проекта с одной стороны носят образовательный и просветительский характер, а с другой - имеют актуализирующую направленность, призванную привлечь внимание к научному потенциалу АН РТ.

В статье рассмотрен подход к созданию информационной системы «Научные школы РТ» по модели виртуального музея-библиотеки, включающего архивные материалы всех видов, в том числе мультимедиа, а также современные информационные интерактивные ресурсы (форумы, видеоконференции, интерактивные модели, обучающие курсы и др.). Информационная система поддержки специализированного виртуального музея-библиотеки (ВМБ) Академии наук РТ позволит реализовать одну из важных и актуальных задач по систематизации, сохранению и пропаганде научных достижений и общественно-значимого наследия ученых АН РТ и доведению до широких масс информации о результатах их научной и общественной деятельности. На сегодняшний день результаты научной деятельности ученых отображаются разрозненно в разных источниках или находятся в труднодоступных личных библиотеках и архивах. Чаще всего исчезают бесследно и безвозвратно после ухода из жизни ученого, и как показывает горький опыт, это не зависит даже от величины его деяний.

Материалы музея-библиотеки позволяют представить наиболее полно и концентрировано деятельность каждого ученого Академии наук, ведущих ученых институтов Академии наук РТ и Академии наук в целом. Научный, общественный и личностный опыт членов Академии окажет несомненное влияние на развитие научно-исследовательской деятельности АН РТ, а также будет способствовать популяризации науки в молодежной среде и сохранению научного наследия для последующих поколений практически навсегда.

## ЛИТЕРАТУРА

1. Максимова Т. Е. Виртуальные музеи как социокультурный феномен: типология и функциональная специфика. Автореф. дисс на соискание ученой степени канд. культурологии. М., 2012. Доступ по <http://cheloveknauka.com/virtualnye-muzei-kak-sotsiokulturnyy-fenomen>)
2. Максимова, Т.Е. Виртуальные музеи: анализ понятия/ Т.Е.Максимова // Вестник Московского государственного университета культуры и искусств. - 2012. - №2. - С. 196 -202.
3. Максимова, Т.Е. Виртуальные музеи: подходы к типологии / Т.Е. Максимова // Вестник Московского государственного университета культуры и искусств. - 2012. - №4. - С.186-190.
4. История информатики в лицах. Электронный ресурс. [http://www.sbras.ru/win/elbib/data/show\\_page.dhtml?77+1213](http://www.sbras.ru/win/elbib/data/show_page.dhtml?77+1213)



# ИНТЕГРИРОВАННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА ТЕМАТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

М.В. Энгель, В.В. Белов

*Институт оптики атмосферы им. В.Е. Зуева СО РАН*

**Введение.** В настоящее время в центрах приема и обработки спутниковых снимков накоплены данные результатов многолетних измерений, полученных различными спутниковыми системами. Для эффективного использования этих данных создаются информационные системы, ориентированные на решение различных прикладных задач. Работы в этом направлении начинались с разработки систем архивации и долговременного хранения данных дистанционного зондирования Земли (ДЗЗ). Следующим этапом развития информационных систем на основе данных такого рода стала организация распределенных систем. В качестве примера реализации этого подхода можно привести систему EOSDIS (The Earth Observing System Data and Information System), созданную на базе архивов спутниковых данных центров NASA, web-сервисы которой предоставляют доступ к тематическим продуктам о параметрах земной поверхности и атмосферы [1].

На основе распределенных многомерных архивов спутниковых данных активно развиваются технологии создания информационных систем, обеспечивающих не только хранение и навигацию, но и инструментарий для анализа и управления данными дистанционного зондирования Земли (ДЗЗ). Примером успешной реализации системы, предназначенной для пространственного, временного и статистического анализа данных различных спутниковых систем, можно назвать портал GIOVANNI [2], работающий на основе информации, хранящейся в архивах сети EOSDIS.

В нашей стране в Институте космических исследований (ИКИ) РАН в конце 90-х годов прошлого столетия была разработана и реализована технология построения системы архивации и долговременного хранения спутниковых данных, позволяющая организовывать автоматически пополняющиеся электронные архивы ДЗЗ. В настоящее время одним из актуальных направлений является создание информационных технологий, обеспечивающих быструю разработку систем тематической обработки и анализа данных ДЗЗ, таких, например, как технология GEOSMIS [3], предназначенная для создания инструментов работы с большими распределенными многомерными архивами спутниковой информации, обеспечивающих не только для поиск, но и задачи анализа и управления данными.

Анализ современных технологий разработки информационных систем для поддержки решения различных прикладных задач на основе ДЗЗ

показывает тенденцию развития таких систем в направлении интеграции уже существующего информационного и алгоритмического обеспечения. Базовой задачей разработки такой системы является выбор информационной модели распределённых данных, обеспечивающей выполнение процедур первичной и тематической обработки данных. Существуют разные подходы к решению этой задачи. Например, для разработки модели распределённых данных и организации их обработки используется парадигма событийно-ориентированного программирования [4]. Авторы работы [5] реализуют программную логику в виде набора сценариев и библиотек.

Нами предложена модель интеграции разнородных распределённых данных ДЗЗ и алгоритмов тематической обработки, основанная на использовании теории автоматов. Реализация предложенной модели показана на примере создания информационной системы атмосферной коррекции спутниковых изображений.

Автоматный подход к построению информационных систем основан на понятии "состояние". Основное свойство состояния заключается в «отделении» будущего от прошлого. Текущее состояние несет в себе всю информацию о прошлом системы, необходимую для определения ее реакции на любое входное воздействие, формируемое в момент времени  $t_0$  [6,7]. При проектировании системы определяются все возможные состояния, в которых может находиться система, и переходы между ними в результате влияния некоторых событий. Из множества возможных состояний выделяются базовые управляющие состояния, которые определяют логику работы системы. На основе множества базовых состояний определяются входной и выходной алфавиты автоматной модели.

Каждый алгоритм обработки, интегрированный в систему, определяется как конечный автомат. Механизм взаимодействия автоматов реализован в виде обмена сообщениями через общую память. Выбор именно этого способа взаимодействия обусловлен асинхронным характером процессов, выполняемых системой и, как следствие, невозможностью синхронизировать работу автоматов между собой. Асинхронность порождена, в первую очередь, возможной недетерминированностью входной информации, процедура получения которой зависит от различных внешних по отношению к системе факторов и может быть определена только в процессе выполнения.

В целом управляющая часть системы представляет собой совокупность автоматов, взаимодействующих между собой через общую память.

Использование автоматного подхода позволяет автоматизировать проведение процедур обработки данных. Вычислительная среда в автоматическом режиме планирует процесс выполнения задачи таким образом, чтобы процесс максимально соответствовал установленному критерию оптимальности. Автоматная модель системы обеспечивает расширяемость, наблюдаемость и сервис-ориентированный характер системы.

**Интеграция распределенных данных на примере информационной системы атмосферной коррекции спутниковых изображений.** При выполнении процедур тематической обработки данных ДЗЗ должен быть решен вопрос о необходимости проведения процедуры коррекции вклада искажающих факторов атмосферы в результаты измерений спутниковых приборов (так называемой процедуры атмосферной коррекции – АК). Наиболее точная коррекция искажающих факторов атмосферы возможна на основе физического подхода, для реализации которого необходимо выполнение следующих условий:

- наличие высокоточной математической модели атмосферы;
- информация о параметрах оптико-метеорологического состояния атмосферы в момент проведения ДЗЗ из космоса.

При выборе возможных источников задания оптико-метеорологических данных об атмосфере необходимо учитывать, что пространственное разрешение данных должно быть сопоставимо с мезомасштабной вариабельностью атмосферных параметров и позволять проводить атмосферную коррекцию всего спутникового изображения, а не отдельных его участков. Таким критериям соответствуют спутниковые и модельные данные. Эффективный подход к проблеме получения оптимального набора данных заключается в комплексном использовании результатов измерений и данных моделирования различных оптико-метеорологических атмосферных параметров.

Постановка задачи предусматривает, что информация может иметь различные характеристики и находится на распределенных ресурсах, количество которых может быть произвольным и меняться в процессе жизнедеятельности системы. Для описания информационной структуры системы используется инфологическая схема распределенных данных, которая включает параметры, обеспечивающие идентификацию распределенных источников и доступ к данным: URL, тип данных, приоритет, сетевой протокол, параметры авторизации. С использованием инфологической схемы реализована настройка данных на работу с алгоритмами тематической обработки (рис. 1).

Выбор спутниковой системы, на данный момент оптимальной для реализации физического подхода к атмосферной коррекции, был предложен и обоснован в работах [8,9]. Здесь в качестве данных о состоянии атмосферы предлагается использовать тематические продукты, полученные на основе измерений спектрорадиометра MODIS [10], а в качестве альтернативного источника метеоинформации – прогностические модели NCEP (National Centers for Environmental Prediction) [11]. Ресурсы для распространения данных обоих типов представлены в Интернете.

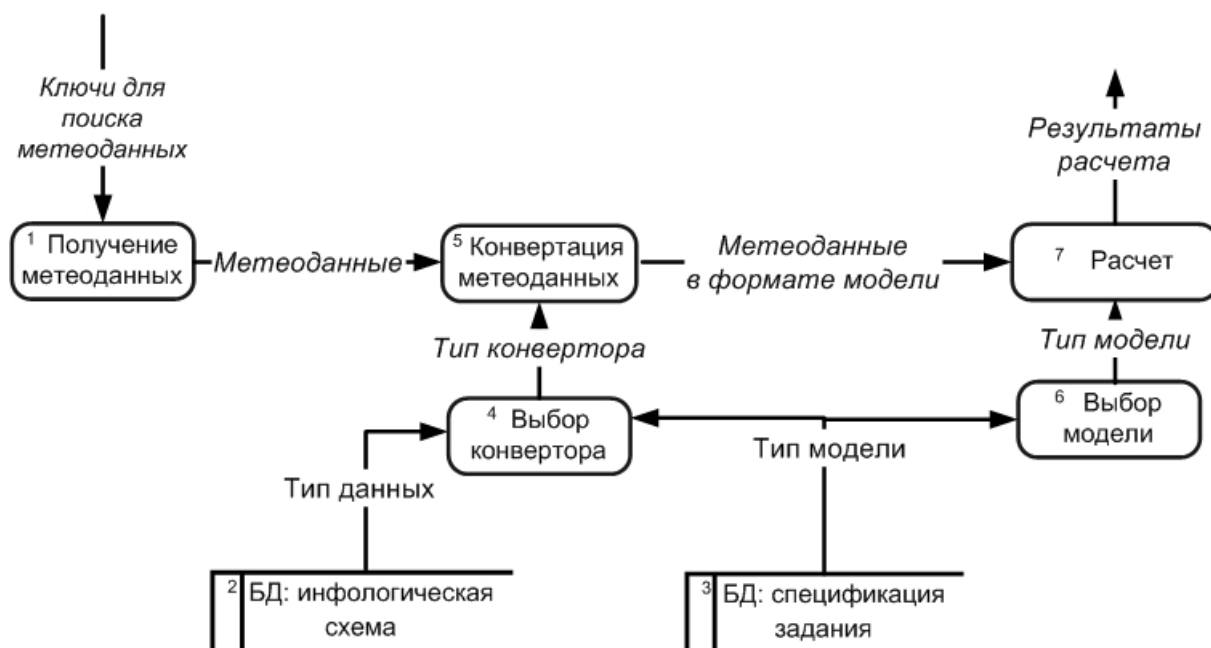


Рис. 1. Схема настройки данных к внутреннему формату модели

В качестве примера сетевого ресурса для распространения спутниковых данных об атмосфере можно назвать LAADS Web (the Level 1 and Atmosphere Archive and Distribution System), Goddard Space Flight Center, NASA [12]. Тематические продукты сенсоров MODIS, содержащие данные об атмосферных параметрах, представлены в архивах LAADS Web за весь период измерений.

Ресурс LAADS Web может быть интегрирован в рамках информационной системы АК в качестве источника информации о параметрах оптико-метеорологического состояния атмосферы, поскольку удовлетворяет следующим требованиям:

- для записи и хранения данных используются стандартные документированные форматы (HDF-EOS, GRIB);
- данные в формате HDF-EOS записаны вместе с метаданными;
- предоставлен доступ к географическим метаданным GeoMeta;
- предоставлен доступ к прогностическим данным, записанным в формате GRIB.

Для реализации процедуры атмосферной коррекции на основе физического подхода выделены следующие управляющие состояния, в которых может находиться система:

- задание параметров расчета;
- определение исходных данных для расчетов на основании заданных параметров;
- получение данных для расчетов;
- оценка релевантности спутниковой информации;
- проведение расчетов;

- формирование и отправка сообщения пользователю;
- нормальное завершение выполнения задания;
- аварийное завершение выполнения задания.

Управляющие состояния и переходы между ними показаны на UML-диаграмме состояний системы (рис. 2).

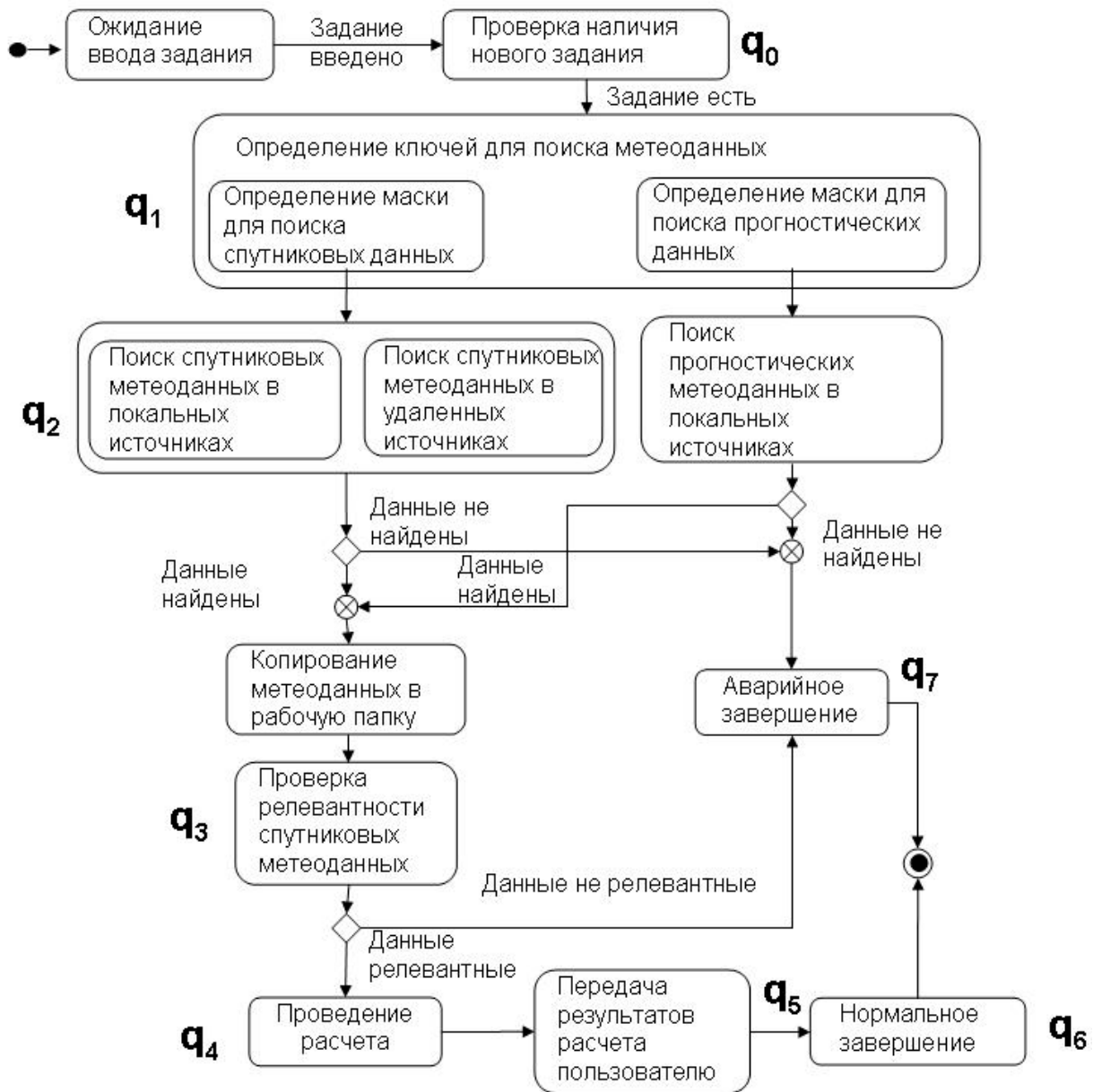


Рис. 2 UML-диаграмма состояний системы

Каждое управляющее состояние представлено в виде пары «автомат и объект управления», где автомат отвечает за логику поведения объекта, а объект выполняет функциональную часть. Таким образом, разработанная система является децентрализованной и представляет собой множество

параллельно работающих процессов, поведение которых описывается конечными автоматами.

**Заключение.** В ИОА СО РАН на основе использования автоматной модели и инфологической схемы данных создан прототип интегрированной информационной системы тематической обработки данных ДЗЗ. На первом этапе разработки в системе реализован алгоритм расчета корректирующих атмосферных поправок. В качестве распределенных источников оптико-метеорологической информации рассматриваются ресурс LAADS Web, а также распределенная инфраструктура ЦКП ДЗЗ СО РАН и электронные архивы спутниковых данных ИОА СО РАН.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-07-06811 А.*

## ЛИТЕРАТУРА

1. Ramapriyan H, Behnke J, Sofinowski E, Lowe D, Esfandiari M. Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS). In: Standard-Based Data and Information Systems for Earth Observation. Springer: 2010.; p. 63-92.
2. The GES-DISC Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni). URL: <http://giovanni.gsfc.nasa.gov/>
3. Толпин В.А, Балашов И.В., Ефремов В.Ю., Лупян Е.А., Прошин А.А., Уваров И.А., Флитман Е.В. Создание интерфейсов для работы с данными современных систем дистанционного мониторинга (система GEOSMIS) // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 3. С. 93-108
4. Д.М. Ермаков, К.С.Емельянов, В.П. Саворский, А.П. Чернушич. Реализация событийно управляемой архитектуры быстрого коллективного доступа к информационным ресурсам ДЗЗ на базе технологии Stream Handler // Современные проблемы дистанционного зондирования Земли из космоса. 2013. Т.10. № 4. С. 118-126
5. И.В. Балашов, О.А. Халикова, М.А. Бурцев, Е.А. Лупян, А.М. Матвеев. Организация автоматического получения наборов информационных продуктов из центров архивации и распространения спутниковых и метеоданных. // Современные проблемы дистанционного зондирования Земли из космоса. 2013. Т.10. № 3. С. 9-20).
6. М.Л. Цетлин. Исследования по теории автоматов и моделированию биологических систем. Наука, 1969. 316 с.
7. Поликарпова Н.И., Шалыто А. А. Автоматное программирование. 2011. СПб.: Питер. 2 издание, 176 с.
8. Афонин С.В., Соломатов Д.В. Методика учета оптико-метеорологического состояния атмосферы для решения задач

- атмосферной коррекции спутниковых ИК-измерений // Оптика атмосферы и океана. 2008. Т. 21. № 2. С. 147-153.
9. Афонин С.В. К вопросу о применимости восстановленных из космоса метеоданных MODIS для атмосферной коррекции спутниковых ИК измерений // Оптика атмосферы и океана. 2010. Т. 23. № 8. С. 684-688.
  10. Level-2 MODIS Atmosphere Products. URL: <http://modis-atmos.gsfc.nasa.gov/products.html>
  11. National Centers for Environmental Prediction. URL: <http://www.ncep.noaa.gov/>
  12. LAADS Web. URL: <https://ladsweb.nascom.nasa.gov/>

