

УДК 004.421.2

ОНТОЛОГИЯ НАУЧНОГО ПРОСТРАНСТВА ИЛИ КАК НАЙТИ ГЕНИЯ

Е.И. Моисеев^{1,2}, А.А. Муромский², Н.П. Тучкова²

¹ Московский государственный университет им. М. В. Ломоносова, Москва, Россия
etc@cs.msu.su

² Вычислительный центр им. А. А. Дородницына РАН, Москва, Россия
tuchkova@ccas.ru

Аннотация

В работе предлагается подход к организации информационного пространства автора, используя тезаурус адресата. Современная задача представляется в том, чтобы грамотно использовать накопленную информацию о публикациях и предоставить специальный сервис для научных работников, который мог бы быть востребован на фоне информационного натиска различных средств коммуникации. В рамках настоящей работы предлагается использовать тезаурус адресата как основу онтологии информационного пространства автора (научного работника), а функциональные средства представить в виде набора сервисов для автора (поиск публикаций, подписка, обмен вторичной информацией, рецензирование, цитирование, оформление в печать, в отчет и др.). На основе описаний предметных областей индивидуумов предполагается построение онтологии научного междисциплинарного сообщества, что, по мнению авторов, позволит не потерять новый результат или открытие в науке, соблюсти приоритеты авторов, встроить новое знание в устоявшуюся систему классических предметных областей.

Ключевые слова: информатизация, онтология, тезаурус адресата, предметная область.

Введение

Научная среда не случайно называется «средой» - местом обитания учёных, педагогов и учеников. Именно в специальных условиях, в окружении коллег развивается учёный какой-нибудь дисциплины, а также сама среда получает развитие и формируется новое знание. Более того, сам факт того, что новое знание появилось, может подтвердиться только в научной среде. Пора учёных одиночек отошла в прошлое, знания интегрированы, объединены многократно в междисциплинарных исследованиях, переплетены интернет-ссылками. Установить первенство в условиях жесткой конкуренции в науке становится сложнее, сеть обеспечивает мгновенное распространение идеи, массив публикаций растет экспоненциально, поисковые системы не всегда имеют доступ к библиотечным системам и базам данных научных журналов. Без онтологии научного пространства трудно определить место научного результата или новой предметной области. Так зачем это множество всего: публикаций, научных исследований и исследователей, когда в громадном информационном пространстве так трудно найти и оценить новое, а прорывные результаты все равно принадлежат единицам, гениям от науки? Тем не менее, все перечисленное создает научную среду, на этом базируется мировое развитие науки и технологии, и именно коллективный труд позволяет проявиться отдельным талантливым учёным.

1 Адресат в научно-информационной среде

Развитие информационных технологий в XXI веке сопровождается увеличением мощностей суперкомпьютеров и гигантским ростом цифрового информационного потока, в кото-

ром специалисты должны усвоить или хотя бы заметить новые факты, утверждения, открытия. Продолжается информатизация образованных слоев общества, и основную роль в этом процессе берет на себя Интернет. Если 5-10 лет назад говорили о Вавилонской башне публикаций, то теперь приходится признать, что Интернету грозит информационное цунами. Этого мнения придерживаются специалисты по проектированию онтологий [1]. Данные исследований показывают, что до 25% бюджета компаний тратятся на проекты интеграции данных и предоставления их пользователям в необходимом и удобном для восприятия виде, а требования к аналитикам данных возросли на 92%.

Отметим (в связи с описанным явлением информатизации общества), что результативность научной работы возможна только при соответствующих условиях. Условия могут быть весьма разнообразны, определяться объективными и субъективными причинами, имея в виду индивидуальность творцов нового знания. В этом плане в истории имеется значительное число примеров, например, легенда об открытии закона всемирного тяготения, открытие гелия, сожжение работы Гельвеция «Об уме» по приговору суда и др. Тем не менее, есть условия общего характера, без которых сегодня каждый активно работающий научный работник вряд ли может обойтись. В данном случае имеется в виду следующее:

- *доступ к научной информации по тематике учёного* (в дальнейшем будем его именовать адресатом) и
- *признание результата научной работы адресата-учёного* как нового знания в масштабах тематики адресата (а именно новизна нового знания адресата в масштабе данной тематики научной среды).

Эти два положения применимы к любому учёному, группе учёных в случае совместной работы, и имеют также свою историю.

Первое положение с незапамятных времен реализовывалось на основе переписки между учёными, например, письма Бернулли, Эйлера, Ферма, Виеты, Паскаля, встречающиеся в некоторых периодических изданиях с небольшими тиражами. В качестве примера сошлемся на издание «Комментарии Петербургской академии наук». За период 1728-1802 гг. вышло в свет 15 томов этого издания. Важность этого издания отмечал Леонард Эйлер в 1743 г. в письме Д. Бернулли [2].

Второе положение имеет весьма большое значение в индивидуальном и в государственном смысле. Последнее отмечалось и отмечается премиями, званиями: например, Д. Гильберт был отмечен как тайный советник [3].

Признание авторства не так уж редко сопровождается явлениями, не имеющими никакого отношения к науке. В качестве примера можно указать на историю метода Лобачевского, что отражает конкуренцию в науке, какова была, есть и усиливается на всех уровнях.

В отсутствие надлежащего финансирования рядовой адресат может не иметь доступа к известным иностранным журналам. Тем не менее, у деятельного адресата в науке есть некоторые возможности, основанные на личной инициативе, целеустремленности и настойчивости, чтобы получить оценку результата своей научной работы. Одна из основных возможностей – работа в Интернет-среде и чаще всего со вторичной информацией [4], в которой можно обнаружить нечто подобное результату адресата, если надлежащей новизны нет.

В данной работе будем опираться на идею создания тезауруса адресата (ТА), (термин «тезаурус адресата (индивидуума)», введен в информатику Ю.А. Шрейдером [5]), как основы для представления предметной области (ПрО) автора в виде онтологии (онтологии адресата).

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

обращаясь к источникам информации, участвуют в процессе информатизации, накопления знаний [6], изменении ТА [7].

В процессе информатизации могут быть случаи:

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

Представим технологию работы с информацией в виде блок-схемы «Адресат в научно-информационной среде».

1.1 Адресат в научно-информационной среде (блок-схема)

На рисунках 1-3 представлена блок-схема типичных событий работы адресата в научно-информационной среде. Описание блок-схемы, связей и переходов, занумерованных от 1 до 42, дано в разделе 1.2.

1.2 Описание блок-схемы «Адресат в научно-информационной среде»

Связи 1, 2, 3, 4, 5, 6, 7, 8 (1-8).

Адресат как пользователь системы Интернет, используя словарь терминов, *формирует запросы*, в которых термины объединяются логическими операторами. На основании составленных запросов адресат образует поисковые образы запросов, учитывая специфику доступа к научной информации. Отметим, что адресат учитывает при составлении запросов тезаурусные отношения между понятиями, обозначаемыми использованной терминологией.

Связи 9, 10, 11, 11а, 11б, 12 (9-12).

Адресат в результате информационного поиска (по тематике адресата и смежным вопросам, см. 6-7) получает информационный массив (документов) (см. 9-10). После этого происходит процесс познания – *анализ полученного информационного массива с точки зрения релевантности* (11, 12) и нерелевантности (11а, 11б).

Напомним, что «*релевантная информация*» – информация, соответствующая тематике запроса; релевантность (relevance) – характеристика степени соответствия. Здесь нет абсолютного критерия (но есть соответствие тому, что смогли найти, нашли то, что есть в сети, а оно может и не соответствовать информационной потребности).

Связи 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 (13-22).

Адресат работает с релевантной информацией, которая с его точки зрения может быть: неполной информацией (13), избыточной информацией (14, 17), полной информацией (15, 19).

В случае *неполной информации* поиск повторяется (поиск соответствующей работы над запросом или поисковым образом, 16-16а, 2-8).

В случае *избыточной информации* осуществляется отбор документов до получения массива полной информации (14-17, 17-18).

Полная информация (15-19) достаточно соответствует тематике запроса (по мнению адресата). Напомним: «*полная информация* – информация, позволяющая с исчерпывающей полнотой ответить на поставленный вопрос». Заметим, что «*полнота информации*» с индивидуальной точки зрения определяется в зависимости от кругозора и знания адресата.

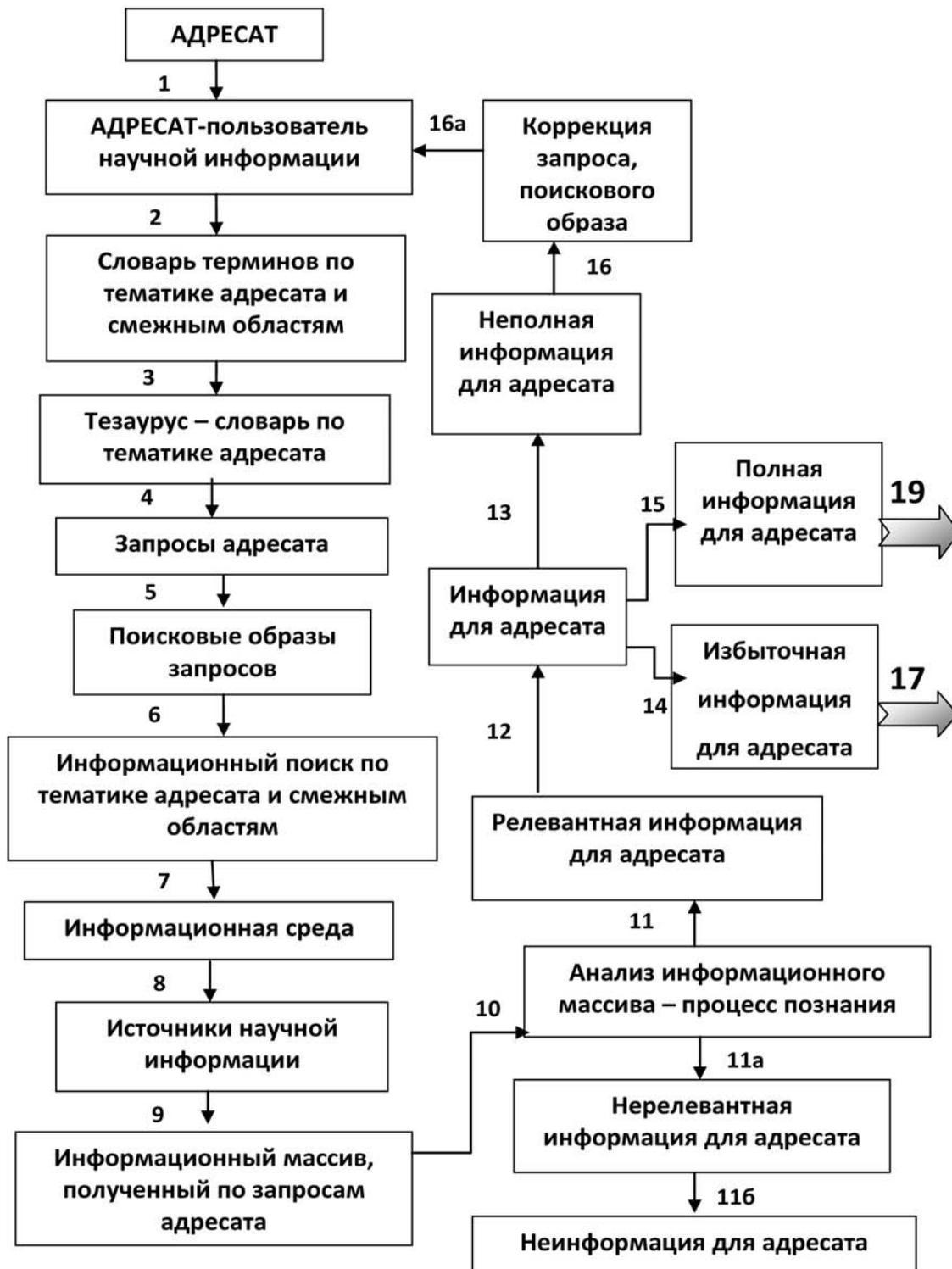


Рисунок 1 - Блок-схема действий адресата в научно-информационной среде (связи 1-19)

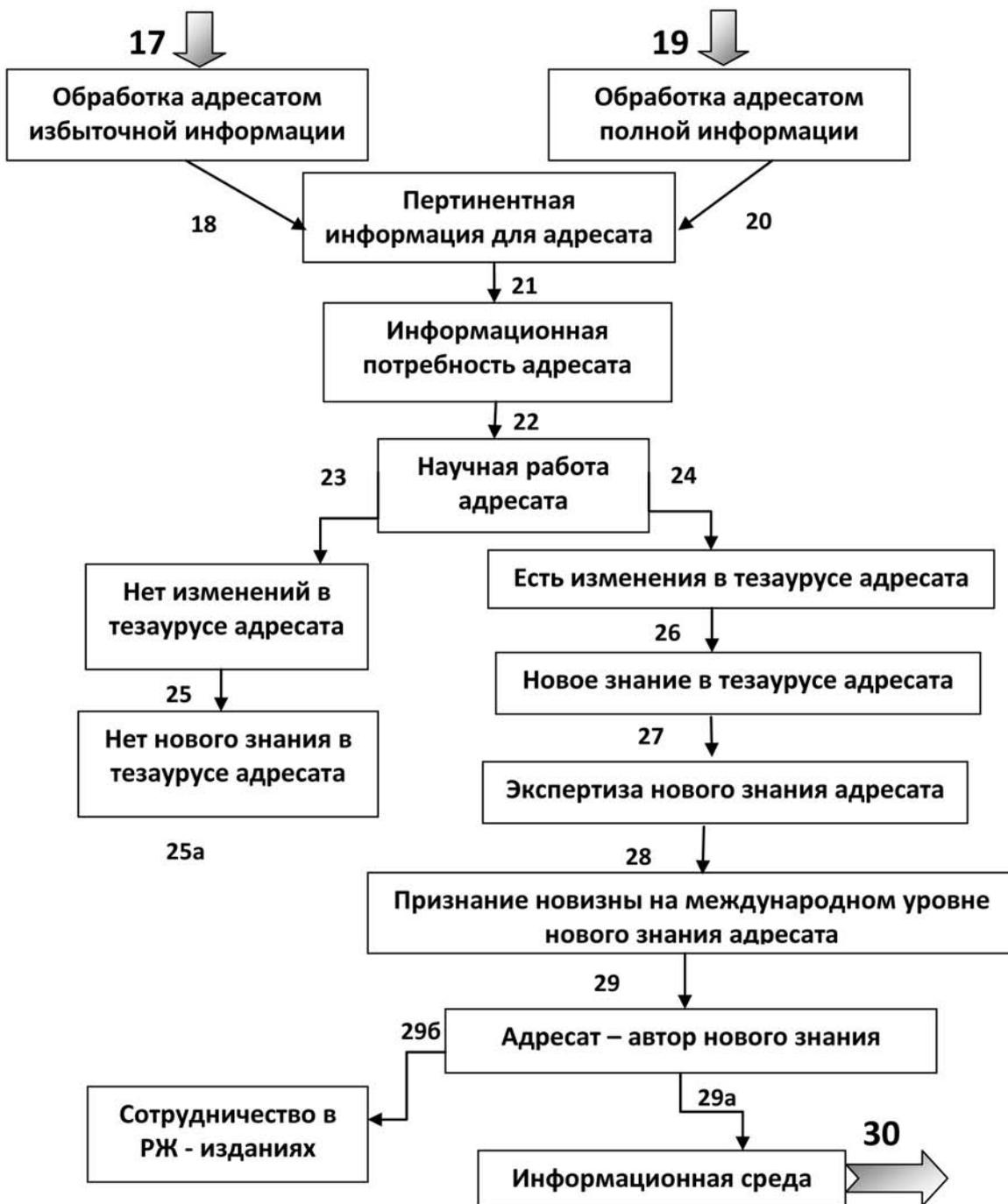


Рисунок 2 - Блок-схема действий адресата в научно-информационной среде (связи 18-30)

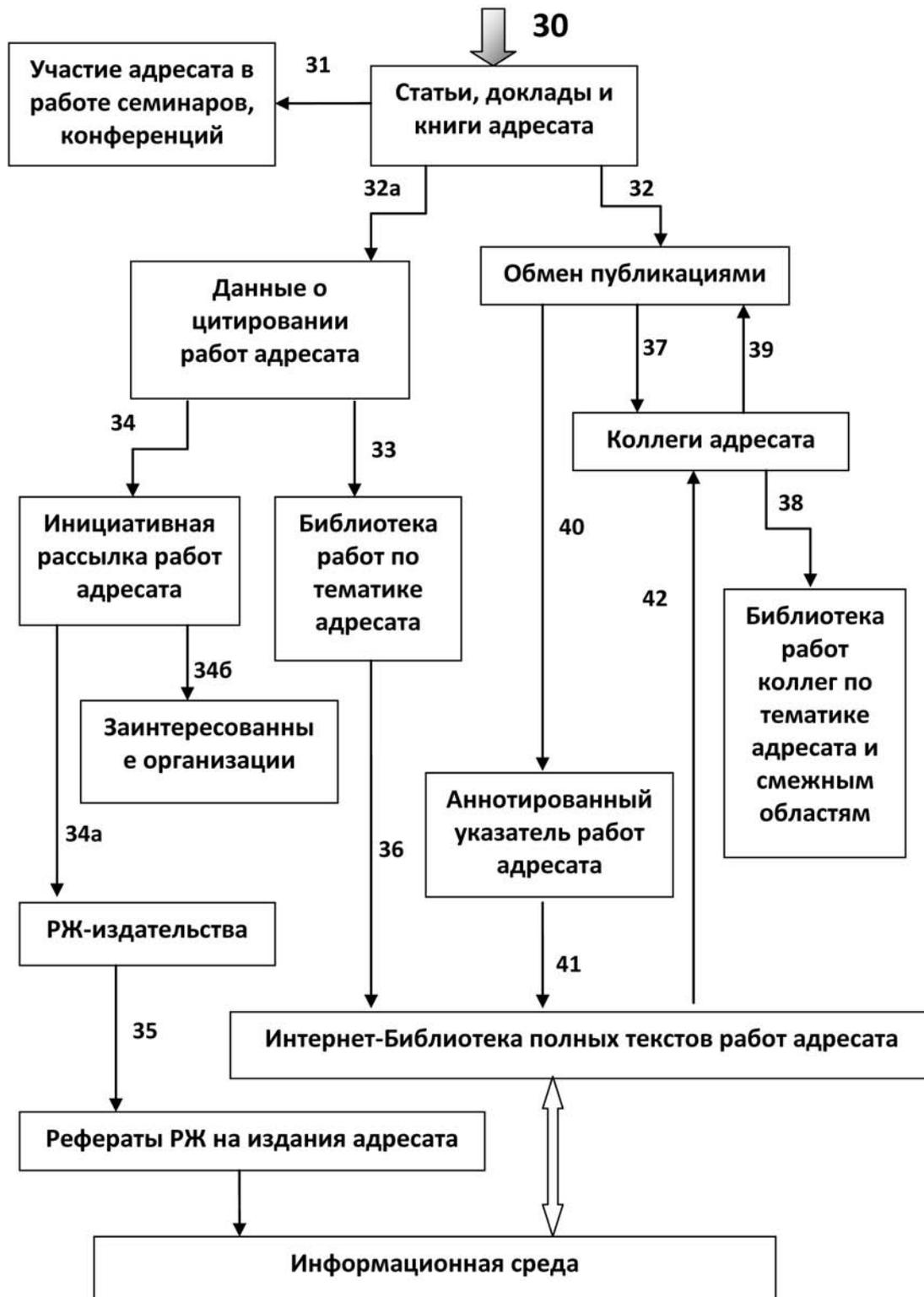


Рисунок 3 - Блок-схема действий адресата в научно-информационной среде (связи 31-42)

Полная информация и обработанная информация (как часть избыточной информации) анализируется с точки зрения пертинентности и информационной потребности адресата (см. 18, 19, 20, 21).

Напомним, что «пертинентность (англ. *pertinence*) – характеристика степени соответствия содержания документа, полученного в результате поиска, информационной потребности адресата (англ. *information demand*)» – свойство отдельного в данном случае лица, адресата, отображающее необходимость получения соответствующей информации. Отметим, что *суждение о полноте информации, информативности документа определяется индивидуально, зависит от эрудиции пользователя, адресата, хотя есть и иное, когда полученная информация бесспорна в рассматриваемой среде пользователей.*

Связи 22, 23, 24, 25, 25а, 26, 27, 28, 29, 29а, 30, 31, 32.

Наконец, *этап научной работы* (22, 23) на предмет определения новизны знания, полученного адресатом, с точки зрения адресата. Здесь используем известное в информатике понятие «тезаурус индивидуума/ адресата».

Если *нет изменения тезауруса адресата*, то информация, полученная адресатом, не является для адресата информацией (информация для адресата), ибо нет нового знания в тезаурусе адресата (23, 25, 25а).

Если *есть изменение в тезаурусе адресата*, то в тезаурусе адресата возникло «*новое знание в тезаурусе адресата*».

В итоге это новое знание может быть:

- новым только в рамках кругозора адресата;
- как новое знание адресата в международном масштабе;
- новым знанием в научно-информационной среде коллег, к которой принадлежит и сам адресат – создатель нового знания.

Естественно, что необходимо признание новизны со стороны коллег упомянутой выше среды (24, 26, 27, 28, 29, 29а, 30, 31, 32, 32а).

Учитывая возможное несовершенство распространения информации, адресат должен проявить некоторую инициативу (см. 34, 34а, 34б), что иногда накладно и не только из-за недостатка средств.

Далее, на блок-схеме представлены составные части библиотеки адресата (33, 38, 42). Библиотека пополняется всеми возможными способами.

На схеме специально отмечены случаи обмена публикациями:

- получение адресатом публикаций от коллег (32, 37, 38);
- получение коллегами публикаций от адресата (39, 40, 41, 42).

Подчеркнем, представленная здесь блок-схема – всего лишь схема. События здесь могут быть в различные периоды (например, 28, 29, 29а).

Иногда какая-либо *информация являлась неинформацией для адресата* (так как не было изменения тезауруса адресата), а спустя какой-то период, ныне, прежняя информация «*стала*» *информацией для адресата* (так как произошло изменение тезауруса адресата). Тезаурус адресата также меняется со временем.

На блок-схеме указаны действия, события, которые типичны для обеспечения нормальной работы адресата. На схеме РЖ – реферативные журналы.

Все разнообразие научной работы невозможно вместить в какую-либо схему без каких-либо ограничений. Здесь отмечены действия желательные. Адресат должен проявлять самостоятельность, которая является одним из основных факторов научной работы, хотя авторитеты есть и будут в любой науке. Как отмечалось еще в 90-е годы специалистами по искусственному интеллекту [8], смысл систематизации информации с использованием онтологического инжиниринга состоит в обеспечении обмена знаниями и только это, в конечном счете,

может являться критерием в развитии этого направления. Желательно, чтобы эти же цели, наряду с признанием собственных результатов, лежали в основе информационного обмена адресата, что частично отображено в блок-схеме.

2 Управление данными о публикациях на основе онтологического подхода

С появлением идеи Semantic Web онтологии играют ключевую роль в моделировании систем управления интеллектуальными данными [10]. В то же время проблема представления ПрО адресата (индивидуума, автора, научного работника) в информационной среде продолжает рассматриваться различными способами, что определяется целевым назначением информационного ресурса. Практически во всех библиотечных, издательских, складских (для книжных магазинов) системах, а также в «профессиональных» соцсетях реализуется некоторое собственное специальное описание информационного окружения автора публикаций. В области научных публикаций - это системы цитирования, например webofknowledge.com, www.scopus.com, elibrary.ru, и социальные научные сети, например www.socionet.ru, nanopub.org, платформа geres.org и др. Однако, как отмечается в публикациях на эту тему [8, 10], нет никаких однозначных описаний ПрО с помощью онтологий, так как любая онтология предполагает развитие, а также зависит от ее использования, цели создания и применения. Более того, существует тенденция расширения онтологий за счет баз данных, отдельных предметных онтологий, включения онтологий смежных областей (импортирование, логическое объединение, добавление описаний) и других информационных ресурсов. Онтология по праву занимает центральное место (или начинку) знаменитого «слоеного пирога» Тима Бёрнерса-Ли [11].

Не оспаривается тот факт, что включение автора в информационную среду (ИС) через ПрО автора способствует естественному объединению авторов соответственно профессиональным интересам в системе научной информации посредством коммуникационных систем. Информация, помещенная в ИС, становится доступной научному сообществу, заинтересованному в разработках определенной ПрО. Пересечение ПрО авторов и авторских коллективов позволяет получить дополнительную информацию об актуальности исследований, приоритетах полученных результатов, формировать ПрО междисциплинарной научной среды на основе соответствующих операций с онтологиями (объединение, включение, наследование).

В данном исследовании онтологии адресата опираемся на множество ключевых слов (КС) из произведений автора. Именно это множество составляет основу для создания ГА и модели ПрО его исследований на основе онтологии ПрО адресата. Под коммуникационной средой понимаем средства взаимодействия персон(ы) в Интернете друг с другом и/или приложениями (программами), средствами Интернет, специально не уточняя платформы.

2.1 Об онтологии предметной области и тезаурусе адресата

Остановимся на варианте описания онтологии ПрО адресата в научной коммуникационной среде применительно к разделам математики и ее приложений. Ограничимся описанием основных классов и подклассов, одним уровнем слотов (свойств) и примерами экземпляров из ПрО обыкновенных дифференциальных уравнений (ОДУ) [12] и ПрО «математическая физика и смежные области» (МФ и СМобл) [13].

Задание основных классов¹, подклассов, слотов

классы:

<словарь>

подклассы: <ключевые слова>

<публикация>

<статья>

<монография>

<тезисы>

<труды конференции>

<препринт>

слоты: <вторичные документы>

<персона>

<автор>

<персональные данные>

<публикации>

<рецензент>

<предметная область>

<место работы>

<контактная информация>

<читатель>

<список запросов>

<идентификатор>

<издание>

<журнал> | <газета> | <сборник> | <книга>

<выходные данные>

<контактная информация >

<импакт-фактор>

<издательство >

<учреждение> | <издательская группа>

< контактная информация >

Под описанием функций персоны (адресата) подразумевается список возможных действий, поддерживаемых системой, содержащей онтологию Про автора, для зарегистрированного пользователя (автора). Схематически они представлены на рисунке 4.

Автор работает с информацией и ему доступны:

- **ввод информации** – создание нового раздела информации;
- **редактирование** существующего раздела информации;
- **подписка на информацию**:
 - по ключевым словам,
 - по авторам,
 - по названию;
- **рецензирование**;
- **научная переписка и совместное редактирование**, «информация» в данном случае:
 - другой адресат,
 - другая предметная область,
 - смежная предметная область,
 - публикации,

¹ Классы, выделены полужирным шрифтом, подклассы и слоты смещены вправо

- вторичная информация о публикациях,
- отчет по публикациям,
- корректура.

Основное назначение предложенного описания ПрО адресата на основе технологии тезаурусного представления знаний – создание условий для поиска контактов и публикаций, подготовки публикаций, рецензий и отчетов по тематике адресата – научного работника. В качестве примера представления таксономических отношений рассмотрим статью лексико-семантического указателя (ЛСУ) по ОДУ и терминологическое обозначение уравнения с частными производными МФ и СМобл.



Рисунок 4 - Реализация базовой функциональности: онтология-персона-публикация

В примерах 1 и 2 приведены онтологические описания соответствующих статей из ПрО ОДУ [12].

Пример 1. Статья ЛСУ (класс СЛОВАРЬ[°], значение слота^{°°}), экземпляр:

[°] ОДУ первого порядка, разрешенное относительно производной, именованное

^{°°} Абеля ОДУ первого рода

^{°°} Абеля ОДУ второго рода

^{°°} Бернулли ОДУ

^{°°} Дарбу ОДУ

^{°°} Муанье ОДУ

^{°°} Пирсона ОДУ

^{°°} Риккати ОДУ обобщенное

^{°°} Риккати ОДУ общее

^{°°} Риккати ОДУ специальное

- Риккати ОДУ каноническое
- Эйлера ОДУ первого порядка
- Якоби ОДУ первого порядка
- Коши задача ОДУ первого порядка, разрешенного относительно производной

Пример 2. Уравнения с частными производными:
класс СЛОВАРЬ°, АВТОР°, ПУБЛИКАЦИЯ°, СЛОТЫ°◦

- Уравнение Лаврентьева-Бицадзе
- $u_{xx} + \text{sign } y \cdot u_{yy} = 0$
- Сабитов К.Б.
- Уравнения математической физики
 - Учебное пособие для вузов
 - М.: Высшая школа.
 - 2003
 - 255

2.2 Классификационные схемы и библиотеки предметной области адресата

Формирование состава ПрО в коммуникационной среде основывается на множестве элементов, входящих в описание публикации этой ПрО (множество вторичных документов, метаописания публикаций), и множестве дополнительной информации, связанной собственно с процессом коммуникации.

Состав ПрО конкретного адресата и смежных областей:

- УДК (UDC)5/54, содержащая деление и соответствующую терминологию по естественным наукам, в том числе по математике, физике, химии и др.;
- рубрикаторы (отечественные/иностраные);
- Mathematics Subject Classification (MSC), классификации, используемые в Zentralblatt и Mathematical Reviews;
- тезаурусы (классификация и терминология, словари контролируемой лексики);
- некоторые классификации в виде отдельных кодов, принятые в РЖ, научных издательствах;
- поисковые образы (запросы), составленные непосредственно адресатом и применяемые им реально (образы содержат как термины, так и коды, и специальные символы, типа «и», «или» и др.);
- вторичная информация (рефераты, аннотации, ключевые слова и др.);
- программы и ссылки на программы (систематически используемое программное обеспечение), обеспечивающие реализацию запросов пользователя – адресата;
- программы и ссылки на программы (программное обеспечение), относящиеся (по мнению адресата) к предметной области;
- первичная информация (сами публикации) адресата;
- первичная информация (сами публикации) и/или вторичные документы, интересные автору.

В процессе взаимодействия с другими авторами и источниками автор с помощью сервисов коммуникационной среды формирует *главный ресурс ПрО*, а именно *словарь терминов ПрО и связанные с ним списки трудов, составляющие множество библиотек авторского ПрО*, а также библиотеки адресата:

- словарь терминов по тематике адресата и смежным вопросам на русском и английском языках (тезаурус);

- библиография работ адресата на русском и английском языках;
- библиотека работ адресата с РЖ – документацией;
- библиотека полных интернет-текстов работ автора-адресата;
- аннотированный указатель работ адресата на русском и английском языках;
- список источников информации по тематике адресата и смежным вопросам в печатных и электронных вариантах;
- «книгообмен», обмен публикациями с коллегами. инициативная рассылка работ адресата и подписка адресата на работы коллег;
- данные о цитировании (рецензировании) автора.

2.3 Реализация системы поддержки предметной области адресата

Отметим некоторые особенности информационных систем, необходимые, на наш взгляд, для плодотворного Интернет-общения адресата:

- реализация поиска публикаций, используя в поисковых запросах вторичную информацию о публикациях, а именно по авторам, ключевым словам, кодам классификаторов (УДК, MSC и др.);
- тематический поиск публикаций (с использованием рубрикаторов);
- поиск публикаций в рамках заданной ПрО (с использованием предметных тезаурусов);
- поиск коллег в заданной ПрО с использованием тех же механизмов систематизации публикаций (вторичные документы, классификаторы, тезаурусы);
- реализация подписки на публикации заданной ПрО (для организации адресной доставки информации);
- приглашение коллег ознакомиться с результатами и публикациями (особенно важно с точки зрения соблюдения приоритетов и поддержания актуальности исследований, рецензирования и т.п.);
- возможность отслеживания цитирования работ автора (требует дополнительных связей с издательскими системами);
- возможность создания личных архивов публикаций с соответствующим сервисом по подготовке, систематизации публикаций для различных целей (отчетов и т.д.).

Сразу заметим, что многое из перечисленного списка реализовано по отдельности в издательских системах, системах подготовки отчетности, в соцсетях и сетях профессиональных сообществ. Отметим российские ресурсы, которые отражают научную профессиональную направленность, такие как www.machinelearning.ru (профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных) [14], www.socionet.ru [15] (в основном отражает социологическую и экономическую тематику, в силу специфики авторов-разработчиков).

В разрабатываемой в рамках данной работы системе функциональные сервисы направлены на организацию информационной поддержки исследований автора по некоторой заданной ПрО. Персональная страница автора формируется на основе авторизованного доступа к системе накопления публикаций. Для работы с публикациями автору предоставляются сервисы: ввода и редактирования списка публикаций, формирования тезауруса, поиска публикаций с указанными ключевыми словами из персонального тезауруса; расширения (объединения) списка ключевых слов за счет списков других авторов, авторизованных в данной коммуникационной среде; поиска коллег по ключевым словам; ввода и редактирования списка цитируемой литературы; обмена списками ключевых слов и списками публикаций с другими авторизованными авторами, услуга «подписка на публикации в системе» с ключевыми словами из заданного списка автора. Поисковые образы (запросы) содержат как термины, так

и коды. Наиболее частные запросы указывают на *актуальность тематики и востребованность отдельных авторов*.

Поиск по ключевым словам позволяет учесть специфику и тематику публикаций, а именно ПрО автора. Основная функция по обмену информацией между авторами реализуется благодаря процедуре контроля ключевых слов. Для этого в системе хранится информация о всех ключевых словах авторизованных пользователей системы и ссылки на их публикации. При пополнении индивидуальных коллекций публикаций и ключевых слов осуществляется рассылка по подписке на публикации с данными ключевыми словами.

Реализация аналогичных проектов уже ведется в европейском научном сообществе, например, немецкий проект ResearchGate (<http://www.researchgate.net/>) или американский проект профессиональных сообществ <http://www.linkedin.com/>. Однако в них в большей степени реализуются задачи социальной сети, а не поддержки научных исследований, хотя отдельные способные участники с помощью этой сети несомненно могут быть найдены.

Заметим, что вопросами *адресной доставки информации, как библиотечной услуги*, занимались и занимаются достаточно давно (режим ИРИ - Избирательного Распространения Информации, рассылка по подписке и пр.). Частичная реализация такой задачи обеспечивается имеющимися в сети электронными библиотечными системами, системами цитирования, административными информационно-библиографическими системами отчетности. Тем не менее, именно *вопросами адресной доставки информации в сети* занимаются в основном коммерческие ресурсы (amazon.com, например), в которых, как правило, мало или совсем отсутствуют научные публикации (исключение составляют учебники, монографии и компьютерные руководства) и издательские группы. Известна и часть открытых библиографических ресурсов в интернете (бесплатных электронных журналов), но они, как правило, не занимаются адресной доставкой информации, вероятно, в силу специфики организации. Создание такого режима в научной коммуникационной среде представляется авторам насущной задачей информационной поддержки научных исследований.

Заключение

Рассмотрены проблемы представления и поиска научной публикации (научного результата) в сети Интернет с использованием ключевых слов из предметной области автора (научного работника, коллектива). Предложено описание тезауруса адресата и его онтологическое представление, а также проанализированы сценарии возможного «поведения» адресата в научной среде при решении проблемы получения новой информации (нового знания адресата). Предполагается, что авторы научных публикаций заинтересованы в распространении собственных научных достижений, а также в подтверждении их новизны и актуальности в масштабах предметной области и науки в целом. На этом основании предполагается проявление активности со стороны авторов в формировании описания предметной области авторских исследований на основе словарей ключевых слов публикаций при предоставлении соответствующих сервисов на домашних страничках авторов. Тем не менее, проблема мотивации существует и предполагается решать ее предоставлением дополнительных информационных услуг активным авторам – участникам информационного обмена.

Нерешенным пока остается вопрос о согласовании различных онтологических описаний. Эта проблема напрямую связана с одним из основных вопросов современных онтологий: проблемой наследования знаний. Новые предметные области необходимо встраивать в старую систему понятий. Хотя вопросы обновления терминологии и признания самих результатов новыми в новых областях знаний кажутся очевидными, на практике всегда встает вопрос о месте новых направлений в сложившейся системе знаний. Дальнейшие исследования дан-

ной проблемы будут направлены создание тезаурусов новых предметных областей и их онтологических описаний.

Благодарности

Работа выполнена при финансовой поддержке РФФИ, проект № 13-07-00334-а.

Список источников

- [1] Data tsunami: are you steering for disaster in a canoe? (Posted on January 24, 2014 by Ontology) <http://www.ontology.com/news-and-media/data-tsunami-are-you-steering-disaster-canoe/> (обращение 14.10.2014).
- [2] **Рыбников, К.А.** История математики. М.: Изд-во МГУ Т. I, 1960. 190 с.; Т. II, 1963, 335 с.
- [3] **Рид, К.**, Гильберт. С приложением обзора Германа Вейля математических трудов Гильберта. Пер. с англ. И.В. Долгачева. Под ред. Р.В. Гамкрелидзе. М.: Наука, 1977. 367 с.
- [4] **Моисеев, Е.И., Муромский, А.А., Тучкова, Н.П.** О представлении и поиске научных результатов современными средствами в электронной информационно-поисковой среде. М.: МАКС Пресс, 2009. 98 с.
- [5] **Шрейдер, Ю.А.** Об одной модели семантической теории информации // Проблемы кибернетики. Вып. 13. М., 1965. – 254 с.
- [6] **Хорошевский, В.Ф.** Пространства знаний в сети Интернет и Semantic Web (Часть 1) // Искусственный интеллект и принятие решений, 2008, №1. С. 80-97.
- [7] **Шрейдер, Ю.А.** Тезаурусы в информатике и теоретической семантике // Научно-техническая информация. Сер. 2, 1971, № 3.
- [8] **Gruber, T.R.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing // In Int. Journal Human-Computer Studies 43, p.907-928. Substantial revision of paper presented at the Int. Workshop on Formal Ontology, March, 1993, Padova, Italy. Available as Technical Report KSL 93-04, Stanford University.
- [9] **Cawkell, T., Garfield, E.** Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar H. Fredriksson (Ed.). IOS Press, 2001. С. 149-160.
- [10] <http://ontology.cim3.net/OntologySummit/2013/> (Обращение 14.10.2014)
- [11] **Berners-Lee, T.J.** The Semantic Web and Research Challenges <http://www.w3.org/2003/Talks/01-sweb-tbl/slide1-0.html> (Обращение 14.10.2014)
- [12] **Моисеев, Е.И., Муромский, А.А., Тучкова, Н.П.** Тезаурус информационно-поисковый по предметной области: обыкновенные дифференциальные уравнения. М.:МАКС Пресс. 2005. – 116 с.
- [13] **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.
- [14] **Рудаков, К.В., Чехович, Ю.В.** Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // ДАН, 2003, Т. 388, № 1 С.33-36.
- [15] **Козаловский, М.Р., Паринов, С.И.** Классификация и использование семантических связей между информационными объектами в научных электронных библиотеках // Информатика и ее применения. 2012. Т. 6. № 3. С. 32-42.

ONTOLOGY OF SCIENTIFIC SPACE OR HOW TO FIND THE GENIUS

E.I. Moiseev^{1,2}, A.A. Muromskiy², N.P. Tuchkova²

¹Lomonosov Moscow State University, faculty of Computational Mathematics and Cybernetics, Moscow, Russia
cmc@cs.msu.su

²Dorodnicyn Computing Centre of RAS, Moscow, Russia
tuchkova@ccas.ru

Abstract

The paper presents a new approach towards the authors informational space organization with the use of an addressee thesaurus. Current task seems to be in implementation of intelligent use of information regarding publications and to provide scientific workers with a special service that would be competitive amongst other means of communication. An approach towards usage of an addressee thesaurus as a theoretical basis for authors information space and a number of services (publications search, subscription, exchange of secondary information, reviewing, citation, publishing preparation and so on) as practical part is described. An ontology for scientific multidisciplinary collaboration is envisioned to be based on subject domains descriptions of different individuals. The authors state that this approach would prevent data loss in scientific discovery, would allow to implement new knowledge into the existing system of scientific domains.

Key words: informatization, ontology, thesaurus of the addressee, subject domain.

References

- [1] Data tsunami: are you steering for disaster in a canoe? (Posted on January 24, 2014 by Ontology) <http://www.ontology.com/news-and-media/data-tsunami-are-you-steering-disaster-canoe/> (Valid on 14.10.2014)
- [2] **Rybnikov, K.A.** Istoria matematiki [The history of mathematics]. M.: MSU published V. I, 1960. 190 p.; V. II, 1963, 335 p. (In Russian).
- [3] **Reid, C.** Hilbert. With an application of Hilbert's mathematical work by Hermann Weyl. Springer-Verlag. Berlin-Heidelberg - New York. 1970 (Russian translation by I.V. Dolgachev. Ed. by R.V. Gamkrelidse). M. Nauka. 1977. 367 p. (In Russian).
- [4] **Moiseev, E.I., Muromskiy, A.A., Tuchkova, N.P.** O predstavlenii i poiske nauchnyh rezultatov sovremennymi sredstvami v informacionno-poiskovoy srede [About modern techniques in search and representation of scientific results within information-search environment]. M.: MAKS Press, 2009. 98 p. (In Russian).
- [5] **Shreider, Yu.A.** Ob odnoy modeli semanticheskoy teorii informacii [About one model of semantic theory of information] // Problemy kibernetiki. Issue 13. M., 1965. (In Russian).
- [6] **Horoshevskiy, V.F.** Prostranstva znaniy v seti Internet i Semantic Web [Knowledge spaces in the Internet and semantic web] (Part 1) // Iskusstvennyy intellekt i prinyatie resheniy, 2008, N 1. P. 80-97. (In Russian).
- [7] **Shreider, Yu.A.** Tezaurusy v informatike i teoreticheskoi semantike [Thesauri in informatics and theoretical semantics] // Nauchno-tehnicheskaya informaciya. Seriya 2, 1971, N 3 (In Russian).
- [8] **Gruber, T.R.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing // In Int. Journal Human-Computer Studies 43, p.907-928. Substantial revision of paper presented at the Int. Workshop on Formal Ontology, March, 1993, Padova, Italy. Available as Technical Report KSL 93-04, Stanford University.
- [9] **Cawkell, T., Garfield, E.** Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar H. Fredriksson (Ed.). IOS Press, 2001. C. 149-160.
- [10] <http://ontolog.cim3.net/OntologySummit/2013/> (Valid on 14.10.2014)
- [11] **Berners-Lee, T.J.** The Semantic Web and Research Challenges <http://www.w3.org/2003/Talks/01-sweb-tbl/slide1-0.html> (Valid on 14.10.2014)
- [12] **Moiseev, E.I., Muromskiy, A.A., Tuchkova, N.P.** Tezaurus informacionno-poiskovy po predmetnoy oblasti: obyknovennyye differencialnye uravneniya [Informational thesaurus for the knowledge domain "Trivial differential equations"]. M.: MAKS Press. 2005. – 116 p. (In Russian).

- [13] **Moiseev, E.I., Muromskiy, A.A., Tuchkova, N.P.** Internet i matematicheskie znaniya: predstavlenie uravneniy matematicheskoy fiziki v informacionno-poiskovoy srede [Internet and mathematic knowledge: presentation of mathematical physics equations in information environment]. M.: MAKS Press. 2008. – 80 с. (In Russian).
- [14] **Rudakov, K.V., Chehovich, Yu.V.** Algebraicheskiy podhod k probleme sinteza obuchemyh algoritmov vydeleniya trendov [Algebraic approach towards the problem of self-teaching trend generation algorithms synthesis] // DAN, 2003, V. 388, N 1 P.33-36. (In Russian).
- [15] **Kogalovskiy, M.R., Parinov, S.I.** Klassifikaciya i ispolzovanie semanticheskikh svyazey mizdu informacionnymi ob'ektami v nauhnnyh elektronnyh bibliotekah [Classification and usage of semantic links between informational objects in scientific e-libraries] // Informatika i eyo primenenie. 2012. V. 6. N 3. P. 32-42. (In Russian).

Сведения об авторах



Моисеев Евгений Иванович, академик РАН, декан, зав. кафедрой высшей математики и кибернетики МГУ им. М.В.Ломоносова, руководитель подразделения Вычислительного центра РАН, главный редактор международного журнала «Интегральные преобразования и специальные функции» (ITSF). Специалист в области функционального анализа и информационных технологий. Окончил физфак МГУ им. М.В.Ломоносова

Moiseev Evgeniy Ivanovich, academician of the Russian Academy of Sciences, dean and head of department of CS faculty of Lomonosov MSU. Head of department of CCAS, editor-in-chief of the international journal ITSF. The expert in the field of the functional analysis and information technologies. Graduated from faculty of physics of Lomonosov MSU.



Муромский Александр Александрович, старший научный сотрудник ВЦ РАН, к.ф.-м.н., окончил механико-математический факультет Московского государственного университета имени М.В.Ломоносова и МВТУ им. Н.Э. Баумана, долгие годы работал в ВИНТИ. Специалист в области математического анализа и информационных технологий.

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



Тучкова Наталья Павловна, старший научный сотрудник ВЦ РАН, к.ф.-м.н., окончила факультет вычислительной математики и кибернетики Московского государственного университета имени М.В.Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

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