

На правах рукописи



**Хайдаров Шамиль Махмутович**

**Модели и методы интеллектуальной обработки  
математических знаний в информационных системах**

Специальность 05.13.11 —  
«Математическое и программное обеспечение вычислительных машин,  
комплексов и компьютерных сетей»

**Автореферат**  
диссертации на соискание учёной степени  
кандидата технических наук

Казань — 2020

Работа выполнена на кафедре компьютерной математики и информатики в Институте математики и механики им. Н.И. Лобачевского «Казанского (Приволжского) федерального университета».

Научный руководитель: д.ф.-м.н., профессор, заслуженный деятель науки РТ  
**Елизаров Александр Михайлович**

Официальные оппоненты: **Сотников Александр Николаевич**,  
доктор физико-математических наук, профессор, заместитель директора по научной работе Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»

**Невзоров Владимир Николаевич**,  
кандидат технических наук, доцент кафедры систем автоматизированного проектирования Казанского национального исследовательского технического университета им. А.Н. Туполева (КНИТУ-КАИ)

Ведущая организация: Федеральный исследовательский центр «Информатика и управление» Российской академии наук

Защита состоится «01» октября 2020 г. в 16 часов на заседании диссертационного совета КФУ.01.04 на базе ФГАОУ ВО «Казанский (Приволжский) федеральный университет» по адресу: 420008, Республика Татарстан, г. Казань, ул. Кремлевская, д. 35.

С диссертацией можно ознакомиться в Научной библиотеке им. Н. И. Лобачевского ФГАОУ ВО «Казанский (Приволжский) федеральный университет» по адресу 420008, Республика Татарстан, г. Казань, ул. Кремлевская, д. 35 и на официальном сайте КФУ <https://kpfu.ru>.

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2020 года.

Ученый секретарь  
диссертационного совета  
КФУ.01.04,  
канд. физ.-мат. наук



Еникеев Арслан Ильясович

## Общая характеристика работы

**Актуальность темы.** С каждым днем объемы цифровых данных возрастают с колоссальной скоростью, а информационно-коммуникационные технологии (ИКТ) используются практически на каждом этапе научно-образовательной деятельности. Формируется новая система научной коммуникации, которая, совершенствуя сетевую инфраструктуру, налаживает новую систему сервисов интеграции научной информации. Большинство электронных научных ресурсов также доступно через интернет, благодаря чему эффективность работы с научным материалом имеет более высокий, недоступный ранее, качественный уровень.

При проведении новых исследований стало возможным использование всего корпуса накопленных научных знаний. Последнее предполагает создание комплекса технологий, обеспечивающих оптимальное управление имеющимися знаниями, организацию эффективного доступа к ним, а также совместное и многократное использование новых видов структур знаний. Формируются электронные библиотеки, осуществляется перевод редакционных процессов на всех этапах издания научных журналов в электронную форму (начиная с этапа верстки выпусков журналов и публикации научных статей и заканчивая этапом их рецензирования). Перевод редакционных процессов в электронную форму и размещение журналов в сети интернет нацелены не только на облегчение/удешевление работ по изданию научных журналов, но и на расширение целевой аудитории, повышение доступности журналов для научного сообщества.

Как известно, большинство современных электронных коллекций научных документов (научные журналы, сборники научных трудов, диссертации, научные отчеты, архивы и др.) представляет собой наборы неструктурированных документов, на базе которых трудно организовать семантический поиск, извлечение метаинформации и различные информационные сервисы. Кроме того, в настоящее время наблюдается значительное увеличение объема данных, включаемых в коллекции, что в свою очередь создает дополнительные трудности при обработке информации. Поэтому в условиях непрерывного роста объемов, а также многообразия информации сейчас активно развиваются новые подходы, инструменты и методы обработки огромных объемов данных.

При управлении электронными научными коллекциями больших данных они в полной мере остаются актуальными, а также появляются новые задачи, в их числе: семантическая разметка, организация поиска, выделение метаданных, формирование тематических кластеров документов, сбор наукометрической информации, подготовка сборников материалов и др. Насущными становятся

проблемы анализа и управления данными в различных областях с интенсивным использованием данных. Часть описанных теоретических проблем решена в рамках настоящей работы.

**Целями** настоящей работы являются исследование, разработка и практическая реализация методов, моделей и комплекса программных инструментов интеллектуальной обработки математических знаний в информационных системах поддержки издательской деятельности, а также развитие программных сред, обеспечивающих поддержку научной деятельности и её интеграцию в мировое информационное пространство.

Основная **задача**, решаемая в рамках настоящей работы, состоит в создании комплекса информационных систем, обеспечивающего реализацию спектра семантических сервисов управления цифровыми коллекциями, научными изданиями и информационную поддержку научно-исследовательской деятельности в области математики. Эта задача разбивается на следующее множество подзадач:

1. трансформация цифровых коллекций математических научных документов в цифровые форматы, допускающие применение методов текстовой аналитики, в частности, выделение метаданных (после соответствующей семантической обработки текстовых массивов) электронных версий сборников трудов математических научных изданий;
2. создание методов консолидации и семантического управления разнородными цифровыми математическими коллекциями, а также их интеграции в мировое научно-образовательное пространство;
3. создание для описания семантики связей информационных объектов соответствующих словарей и использование существующих онтологий связей (SPAR, SKOS, CERIF, DoCO и др.);
4. разработка алгоритмов классификации физико-математических документов, а также расширение функциональных возможностей управления данными в рамках издательской информационной системы Open Journal Systems, включающих технологию подбора индексов научно-тематических классификаторов (например, УДК, MSC) и методы автоматизации экспертной оценки научного рецензирования;
5. демонстрация возможностей программной реализации предложенных моделей, методов и алгоритмов.

**Объектом исследования диссертации** являются информационные издательские системы в аспекте автоматизации происходящих в них процессов.

В качестве **предмета исследования** в работе выступают формирование комплекса методов и моделей интеллектуальной обработки математических знаний, а также их программная реализация.

**Методология и методы исследования.** В работе применяются методы анализа текстов на естественном языке, методы классификации, методы программной инженерии и методы объектно-ориентированного программирования.

Теоретической и методологической основой диссертационной работы являются разработки отечественных и зарубежных ученых в области информационных технологий, интеграции данных, анализа естественного языка, информационного поиска и машинного обучения, которые представлены в исследовательских работах отечественных и зарубежных ученых: М. Kohlhase, С.Д. Manning, G.S. Ingersoll, S. Peroni, М.Р. Коголовский, А.М. Елизаров, Е.К. Липачёв, Д.Д. Голомазов, J. Chen, F. Ricci, И.А. Резников.

**Научная новизна** работы связана с развитием методов обработки математических знаний и подходов к управлению большими коллекциями цифровых математических документов, основанных на семантических методах, а также созданием новых программных решений и методов построения и систематизации рекомендательных систем. Создание моделей управления математическим содержимым позволит реализовать их на практике, с учетом всех требований, предъявляемых к нему. Исследованные алгоритмы работы с цифровыми физико-математическими документами позволяют применять их как при реализации поставленных задач в издательской информационной системе OJS, так и при развертывании других систем.

**Практическая значимость** исследования заключается в том, что оно вносит вклад в дальнейшее развитие моделей и методов управления математическим контентом. Полученные в ходе исследования результаты работы могут быть применены в информационных системах поддержки издательской деятельности. Материалы исследования или отдельные ее части найдут применение разработчиками программного обеспечения при создании собственных программных инструментов для построения сервисов анализа и интеграции научных документов.

#### **Основные положения, выносимые на защиту:**

1. метод конвертации математических документов в формате OpenXML в  $\text{\TeX}$ -нотацию.
2. методы семантического анализа цифровых научных документов.
3. прототип программного комплекса обработки больших коллекций физико-математических документов.
4. рекомендательная система автоматизированного подбора индексов Универсальной десятичной классификации (УДК) для физико-математических документов.
5. рекомендательная система подбора рецензентов для научных документов в информационных системах.

**Соответствие научной специальности.** Содержание работы соответствует паспорту научной специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», пункты 2, 3, 4 и 9. Сопоставление приведено в таблице 1.

**Апробация работы.** Модели и алгоритмы, предложенные в диссертационной работе, обоснованы корректным применением указанных выше методов

Таблица 1 — Сопоставление направлению и результаты исследований, предусмотренных специальностью 05.13.11

<b>Направление</b>	<b>Результат работы</b>
2. Языки программирования и системы программирования, семантика программ	Разработана семантическая модель организации хранения знаний.
3. Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем	Разработан инструмент конвертации документов в формате OpenXML в TeX. Разработаны инструменты интеграции метаданных научных публикаций в информационные системы, такие как OJS и eLibrary. Разработана рекомендательная система подбора рецензентов для научных документов в информационных издательских системах.
4. Системы управления базами данных и знаний	Разработана система хранения обучающей выборки для рекомендательных систем подбора индексов УДК и экспертов.
9. Модели, методы, алгоритмы и программная инфраструктура для организации глобально распределенной обработки данных.	Разработан программный комплекс семантической обработки физико-математических документов.

исследования. Практическая апробация результатов работы подтвердила эффективность предложенных моделей и алгоритмов.

Результаты работы докладывались на международных и всероссийских конференциях: Международная научная конференция «Аналитика и управление данными в областях с интенсивным использованием данных» DAMDID-2016 (Москва, 2016), XIV и XV Международная конференция по компьютерной и когнитивной лингвистике TEL (Казань, 2016, 2018), Всероссийская научная конференция «Интернет и современное общество» (IMS-2016, Санкт-Петербург, 2016), VIII и IX Международные научно-практические конференции «Электронная Казань» (Казань, 2016, 2017), Вторая Российско-Тихоокеанская Конференция по Компьютерным Технологиям и Приложениям RPC-2017 (Владивосток, 2017), XX и XXI Всероссийские научные конференции «Научный сервис в сети Интернет» (Новороссийск, 2018, 2019), VI и VII международные научно-практические конференции «Информационные технологии в образовании и науке» (ИТОН, Казань, 2017, 2018), Научно-техническая конференция по итогам совместного конкурса фундаментальных исследований РФФИ-РТ в

2018 году (Казань, 2018), а также на различных городских научных семинарах, итоговых конференциях Казанского федерального университета (Казань, 2015–2019).

**Личный вклад.** Все выносимые на защиту результаты получены соискателем лично. В опубликованных совместных работах постановка задачи осуществлялась совместными усилиями соавторов при непосредственном участии соискателя.

**Публикации.** Основные результаты по теме диссертации изложены в 27 печатных изданиях, 5 из которых изданы в периодических научных журналах и сборниках, индексируемых Web of Science и Scopus, 6 – свидетельств о регистрации программ ЭВМ.

## Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируются цели, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы.

**Первая глава** посвящена обзору и исследованию существующих информационных систем поддержки издательской деятельности. Описана модель цифрового издательского процесса. В издательской деятельности использование информационно-коммуникационных технологий позволило передовым современным издательствам не только наладить опережающий выпуск электронных версий научных изданий, но и предоставить авторам, читателям, редакционным коллегиям и редакциям новые сервисы для работы с информацией.

Важным компонентом современных информационных систем управления научными журналами являются сервисы, регулирующие процесс рецензирования и обеспечивающие коллективное редактирование электронных документов. Системы такого типа должны предоставлять такие редакционные сервисы, как классификация, аннотирование, выделение метаданных, публикация, долгосрочное хранение, конвертирование, распространение, синдикация, статистика использования, харвестинг, объединение в коллекцию, взаимодействие с институциональными репозиториями, контроль доступа, подписка, рассылка уведомлений, новые поступления. Помимо удаленного представления статей в научный журнал и их дальнейшей обработки для окончательной публикации информационные журнальные системы обеспечивают доступ к сформированному контенту и расширенный поиск (по автору, названию статьи, ключевым словам и др.) в соответствующих электронных коллекциях, т. е. в полном объеме реализуют функциональные возможности, присущие цифровым (электронным) библиотекам. С этой точки зрения, электронный научный журнал

можно рассматривать как научную электронную библиотеку (ЭБ), оперирующую статьями журнала как информационными объектами.

Проведенный анализ проектов создания системы управления электронными научными журналами позволил сформулировать следующие выводы:

1. практически все системы, связанные с электронными журналами и электронными издательствами (OJS, ePubTK, DPubS, Ambra), были созданы в период 2004–2008 гг. и разрабатывались для обеспечения функционирования конкретных электронных изданий; это привело к существенным различиям как в архитектуре систем, так и в их функциональных возможностях;
2. не существует универсальной модели системы управления электронным журналом с описанием конкретных требований и сервисов; разработчики таких систем часто брали за основу опыт создания конкретных систем управления электронными библиотеками и не использовали в полном объеме результаты, достигнутые в области ЭБ;
3. практически все проекты создания систем управления электронными научными журналами поддерживают общепринятые стандарты в области интеграции и обмена данными;
4. на текущий момент большинство проектов не получило дальнейшего существенного развития; исключением является активно развивающийся проект – Open Journal Systems.

Для обеспечения хранения и поддержки жизненного цикла как отдельных статей, так и журналов в целом, целесообразно использовать платформу OJS в качестве ядра системы управления электронными научными журналами.

Во второй главе рассмотрены методы управления математическим контентом. Как известно, управление контентом включает обработку, хранение, отображение и поиск информации. Особенность управления математическими данными, прежде всего, заключается в структурированности как текста (определения, теоремы, доказательства и т. д.), так и самих формул (вложенности и др.). Приведен обзор различных методов хранения физико-математического контента (нотация  $\TeX$ , офисные пакеты, формат OpenXML). Описаны традиционные форматы представления физико-математического контента в информационных системах.

При условии существования различных форматов хранения математического контента появляется необходимость взаимодействия между ними, а именно, конвертации. Преобразование данных становится необходимым инструментом, позволяющим систематизировать накопленные знания в различных форматах, в том числе, при работе издательских информационных систем. В данной главе предложен вариант преобразования данных из одного формата в другой – представлен метод конвертации документов в формате OpenXML в  $\TeX$ -нотацию. Так как формат OpenXML является частным случаем XML, к нему можно применить язык преобразований XSLT.

Конвертер поддерживает основное форматирование текста, формулы: все математические функции, поддерживаемые Word-ом (дроби, радикалы, тригонометрические функции, пределы и пр.), операторы (интегралы, ряды и пр.), матрицы и системы уравнений, диакритические знаки, греческий алфавит и буквоподобные символы и др. Нельзя не отметить, что  $\TeX$  относительно кодировки Unicode является беднее в плане количества используемых символов. Так как OpenXML использует Unicode, как следствие, есть целые группы символов, присутствующих в формате docx, но отсутствующих в  $\TeX$ -е. Таким образом, необходимо в  $\TeX$  подключить дополнительные пакеты.

Такой подход к конвертированию данных позволяет использовать его на многих платформах с поддержкой XSLT. Все современные веб-серверы имеют такую поддержку. При использовании на персональном компьютере будет достаточно написать небольшое приложение или запустить с помощью любого XLST-процессора (например, SAXON).

В данной главе рассмотрены основные форматы хранения математического контента в издательских информационных системах. Также предложен вариант преобразования данных из одного формата в другой. Преобразование данных становится необходимым инструментом, позволяющим систематизировать накопленные знания в различных форматах, в том числе, при работе издательских информационных систем.

Семантическому анализу цифровых научных документов посвящена **третья глава**. В ней рассмотрены модели и методы выделения метаданных статей на основе структурного анализа, а также варианты организации электронного хранилища, в том числе, с использованием форматов библиографических и реферативных баз цитирования. Также в главе приведен обзор систем для описания структуры документа на основе онтологий.

Решена задача выделения метаданных из неструктурированного текста в формате OpenXML и дальнейшее его преобразование, в том числе, интеграция в информационную систему OJS.

Чтобы научный документ был машиночитаемым, каждый его элемент должен быть размечен специальными метками, которые формируют так называемые метаданные документа. Как правило, любая научная публикация обязательно включает ее библиографическое описание (авторы, название, источник (например, журнал), год издания, том, номер, начальная и конечная страницы), авторское резюме (аннотация, реферат) и ключевые слова. Кроме того, в публикациях могут быть указаны названия и места расположения организаций, от имени которых авторы представили свои материалы (аффилиация); фамилии ученых (членов редколлегии), представивших статью к публикации, дата подачи статьи и другая информация. Отметим, что для размещения в международных информационно-аналитических системах неанглоязычных научных материалов необходимо дополнительно включать в блок метаданных список авторов, название работы, аннотацию, аффилиацию и ключевые слова на английском языке, а список литературы привести в транслитерации.

Для автоматизированного формирования блока метаданных сегодня с успехом могут быть использованы современные информационные технологии. Машиноориентированная обработка электронных коллекций позволяет создать такую семантическую разметку. Коллекция разбивается на классы сходных по структуре документов, для каждого класса производится преобразование документов к семантическому представлению. Особенным методом выделения информационных блоков является технология регулярных выражений, шаблонов используемых для сопоставления последовательностей символов в строках. С помощью набора паттернов регулярных выражений, специфичных для каждого класса документов, производится выделение информационных блоков. Используя регулярные выражения, можно указать правила для множества возможных строк, которые нужно проверить; это множество может содержать строковые фразы, или адреса электронной почты, или TeX-команды и т.п.

Для извлечения метаданных статьи по характерным признакам определяются правила выделения блоков статьи. К таким признакам относятся стилевое оформление статей (шрифт, размер шрифта, выделение и т. д.). Качество извлечения метаданных помогают повысить как такие дополнительные признаки, как положение в тексте, так и шрифтовое оформление. Для каждого блока составляется список специальных слов, которые показывают, что строка или ее часть вероятно принадлежит искомому блоку метаданных. В группу таких слов для аннотации статьи, например, входит «Abstract» или «Аннотация» и т.п. Для извлечения заголовка рассматриваемым методом также используется информация о том, что он может располагаться в начале статьи, выравниваться по центру, а также имеет более крупный и полужирный шрифт.

**ТЕОРИЯ ВАРИАЦИОННЫХ ОБРАТНЫХ КРАЕВЫХ ЗАДАЧ АЭРОГИДРОДИНАМИКИ: СОВРЕМЕННОЕ СОСТОЯНИЕ, ПРИЛОЖЕНИЯ, ПЕРСПЕКТИВЫ РАЗВИТИЯ**

**Д.М. Елизаров**

*Казанский (Приволжский) федеральный университет*

**Аннотация.** Вариационные обратные задачи аэродинамики (ОВЗА) решаются одним из методов оптимизации аэродинамических и гидродинамических форм, в частности, они сводятся к поиску ответа на вопрос, какому максимальной подвинутой спуту можно получить на профиле дна и каков будет профиль, обладающий оптимальными аэродинамическими характеристиками. В рамках классических методов механики жидкости и газа в математическом плане эти задачи сводятся к вариационным краевым задачам для аналогичных функций.

Представлены новые результаты теории вариационных ОВЗА, в том числе, близкие к окончательным, описаны приложения в гидродинамике и теории фильтрации, охарактеризованы перспективы развития.

Работа выполнена при поддержке грантов РФФИ №№ 15-07-05380, 15-47-02343.

**ОСНОВНЫЕ РЕЗУЛЬТАТЫ ТЕОРИИ ВАРИАЦИОННЫХ ОВЗА**

Одним из первых примеров вариационной ОВЗА служат задачи максимизации подъемной силы дуги заданной длины и ограниченной кривизны при безразличном ее обтекании потоком идеальной несжимаемой жидкости (ИЖК). Ее точное решение получено в [1] – доказано, что оптимальная дуга будет окружностью. К названному классу относятся также обтекание обратных краевых задач аэродинамической теории фильтрации, а также экстремальные задачи теории струй.

Пронисные буквы  
Полужирный шрифт  
Расположен в начале текста

Имеет вид И.О. Фамилия  
Студент за названием

Курска  
Расположен после авторов

Размер шрифта: 9pt  
Уникальный вид отрывка

Минимален по словам  
«Аннотация»  
Размер шрифта: 9pt

Рис. 1 – Характерные признаки метаданных в статье, по которым их можно извлечь

В случае поиска авторов учитывается тот факт, что обычно указание на них расположено после заголовка статьи, по центру и удовлетворяет одному из регулярных выражений, описывающих возможное представление списка имен

авторов. Например, имя автора может содержать либо имя и фамилию, либо инициалы и фамилию. Предполагается, что в списке авторов не могут содержаться никакие другие символы, кроме букв, запятых, точек и дефисов.

В электронных научных коллекциях, представленных в офисных форматах (.doc, .docx и др.), а также .pdf, семантическая разметка в тексте отсутствует. Она может быть представлена в виде метаописания документа, однако, как правило, они ничего не имеют общего с содержимым документа. Тем не менее, выполнить такую разметку можно в автоматическом режиме на основе информации о структурном строении каждого документа и особенностях его форматирования.

Для непосредственной работы с форматом, расширяющим язык разметки XML (например, OpenXML, HTML), можно применить технологии семантического веба, такие, как XSLT и XPath/XQuery. Также можно открыть документ при помощи объектной модели, так, например, на языке программирования PHP существуют различные наборы инструментов для работы с XML документами, наиболее распространенные из которых, являются SimpleXML и DOM API. Так, в случае с форматом OpenXML, первым делом потребуется извлечь из документа .docx, который по сути является ZIP-контейнером, файл document.xml.

В файле document.xml содержится основная структура документа, три основных тега в которых содержится текст документа это: «w:p» – для абзаца, «w:t» – выделенная область и «w:t» – текст. Первые два тега могут содержать информацию о форматировании (содержатся в тегах «w:pPr» и «w:tPr» соответственно), причем могут быть как в виде отдельных тегов внутри документа, так и в виде ссылок на примененные стили, или оба сразу. В таком случае сперва применяется форматирование из стилей, а затем указанные в виде тегов. Сами ссылки на стили хранятся в отдельном файле styles.xml – возникает проблема, при которой одни элементы текста в документе выделены непосредственно тегами из файла document.xml, а другие имеют лишь информацию о названии примененного стиля, поэтому, для возможности полноценного использования информации о форматировании текстовых блоков в документе, потребуется предварительное слияние этих файлов. Ввиду того, что формат также хранит историю изменений версий документа, может возникнуть ситуация, при которой слова в тексте могут быть разделены на отдельные теги «w:t», здесь также потребуется предварительное их объединение с соседними, если их стили совпадают. При этом в итоге этих слияний визуально документ не должен измениться.

На основе полученных данных был разработан программный комплекс выделения метаданных из документов в формате OpenXML. Данный программный комплекс позволяет обработать коллекцию физико-математическую документов в представленном формате, проверить на его соответствие, выделить все необходимые метаданные и производить дальнейшую трансформацию в различные форматы, в том числе создание новых оригинал-макетов научного



пытается присвоить метку объекту. Для этого классификатор пользуется знаниями, полученными при изучении примеров уже помеченных объектов. Обучающие данные служат источником априорных знаний, на основе которых классификатор принимает решения об объектах, которых раньше не видел. Ее задача – отнести объект к некоторой категории.

Точность классификатора определяется признаками, на которых он обучен, а также качеством и количеством предъявленных во время обучения примеров. Не имея достаточного количества примеров, классификатор не сможет установить связь между признаками и категориями, поэтому в процессе обучения будут сделаны неправильные предположения о связях. Такой классификатор не сможет провести различие между категориями или предположит, что признак соотносится с некоторой категорией, хотя на самом деле это не так.

Алгоритмы классификации обучаются на примерах, а обучающие данные разбиты на классы вручную или с помощью какого-то автоматизированного процесса. Анализируя связи между признаками и классами, алгоритм учится понимать, какие признаки существенны для правильного присвоения метки, а какие несут несущественную или уводящую в сторону информацию. Результатом процесса обучения является модель, которая впоследствии используется для классификации ранее непомеченных объектов. Классификатор выделяет признаки объектов, подлежащих классификации, и с помощью модели определяет наилучшие метки для каждого объекта. В зависимости от алгоритма классификации может порождаться одна или несколько меток, снабженных оценкой, или вероятностью, благодаря чему метки можно ранжировать.

Для построения рекомендательной системы были применены алгоритмы классификации документов, основанные на свойствах векторной модели. В этой модели документы представляются в виде векторов  $\bar{V}(d_i)$ , компонентами которых являются веса терминов, где под документами мы подразумеваем индексы-классификаторы.

Существуют различные схемы взвешивания терминов, одной из которых является статическая мера

$$tf-idf(t,d,D) = tf(t,d) \times idf(t,D),$$

где  $D$  – множество документов,  $d$  – документ,  $t$  – термин. Известно несколько модификаций функций  $tf(t,d)$  и  $idf(t,D)$ . Например,

$$tf(t,d) = \frac{n_{t,d}}{n_d},$$

$$idf(t,D) = \log \left( \frac{|D|}{|\{d_i \in D | t \in d_i\}|} \right),$$

где  $n_{t,d}$  – количество вхождений термина  $t$  в документ  $d$ ,  $n_d$  – общее количество слов в документе  $d$ ,  $|D|$  – количество документов в множестве,  $|\{d_i \in D | t \in$

$d_i\}$  – количество документов, в которых встречается термин  $t$ ;

$$tf-idf(t,d,D) = \left( \sum_{t \in d} \begin{matrix} 1 & \text{if } t \in d \\ 0 & \text{else} \end{matrix} \right) \times \log \left( \frac{N-n}{n} \right),$$

где  $N$  – количество документов в множестве,  $n$  – количество документов, в которых встречается термин  $t$ ;

$$tfidf(t,d,D) = tf(t,d) \max \left\{ 0, \log \left( \frac{N-DF(t)}{DF(t)} \right) \right\},$$

где  $DF(t)$  – число документов в коллекции, где встречается термин.

Для оценки сходства между документами, представленными в виде векторов, мы применяем косинусную меру

$$sim(d_1,d_2) = \frac{(\bar{V}(d_1), \bar{V}(d_2))}{\|\bar{V}(d_1)\| \|\bar{V}(d_2)\|},$$

а также функцию ранжирования Окари BM25

$$score(d,Q) = \sum_{j=1}^n idf(t_j) \frac{f(t_j,d)(k_1+1)}{f(t_j,d) + k_1 \left( 1 - b + b \left( \frac{|d|}{avgdl} \right) \right)}.$$

Таким образом вычисляются расстояния между предъявленным новым документом и всеми ранее классифицированными и на этой основе производится его классификация. В этом контексте некатегоризованный документ становится запросом, по которому отбираются классифицированные документы или документы, репрезентативные для содержимого каждой категории.

Для формирования рекомендаций, в первую очередь система должна иметь базу терминов по каждому разделу УДК, на основании которой система будет проводить анализ. Этот этап называют этапом обучения классификатора. Для обучения системы был выбран архив физико-математических публикаций Общероссийского математического портала Math-Net.Ru. На портале размещено более 200 тысяч статей в 57 тысячах из которых указан индекс УДК. В автоматическом режиме, были извлечены термины из названия, списка ключевых слов и аннотации, приведенных в документах.

Был использован подход, называемый мешком слов, заключается в том, чтобы рассматривать документ просто как множество слов. Каждое встречающееся в документе слово считается признаком, и этим признакам присваиваются веса в соответствии с частотой вхождения. Для вычисления веса каждого слова используется схема взвешивания  $tf-idf$ , так что важность встречающихся в документе слов зависит от того, насколько часто слово встречается во всем обучающем корпусе.

Для пополнения словаря эти статьи были обработаны, в результате чего были выделены метаданные (название, ключевые слова и аннотация), из которых были извлечены термины. Полученные данные были включены в словарь, который имеет структуру инвертированного индекса в виде семантического представления и является частью создаваемой рекомендующей системы.

В отличие от стандартной векторной модели, здесь документы группируются по указанному индексу УДК, поэтому в качестве документа  $d$ , в методе tf-idf является группа документов. В сформированную обучающую выборку попало 1516 индексов УДК из категории 5 – Математика, и более 28 термов.

Рекомендательная система была реализована на языке программирования PHP и может быть использована, например, в издательских информационных системах для проверки правильности выбора авторами классификатора. Кроме того, система может быть использована в цифровых библиотеках, добавляя дополнительный элемент метаданных для публикаций, у которых не указан классификатор.

Алгоритм формирования рекомендаций по подбору рецензентов состоит из двух частей. Первая часть отвечает за создание и пополнение базы экспертов в информационной системе журнала. Вторая – это блок подбора экспертов для очередной рассматриваемой статьи.

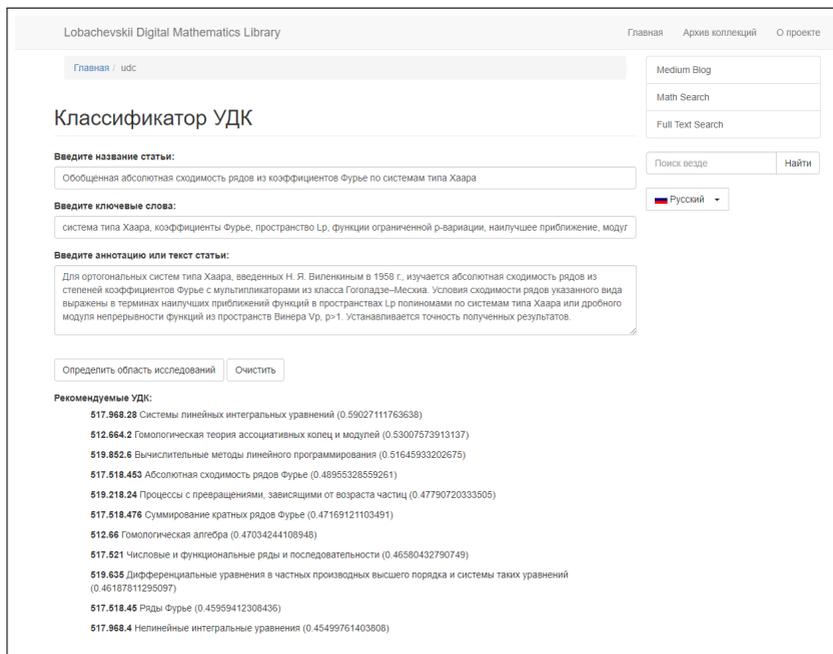


Рис. 3 — Иллюстрация работы прототипа рекомендательной системы

Стандартный функционал платформы OJS позволяет указывать у авторов только ORCID. Система OJS предоставляет возможности изменения своего функционала путем добавления модулей (плагинов), составленных по определенным правилам. Поскольку система написана на языке PHP и является открытой, имеется возможность внести изменения в функционал системы без специального инструментария OJS.

Первоначально автоматизированным поиском в реферативных и библиографических базах данных просматриваются статьи автора, включенного в состав экспертов, в которых имеются коды MSC. В нашем случае были использованы такие базы цитирования как Mathematical Reviews Database (см. <https://mathscinet.ams.org/mathscinet/help/about.html?version=2>) и Zentralblatt MATH (zbMATH, см. <https://zbmath.org>). Для этого в таблицу 'user\_settings' были добавлены два новых поля 'mathscinet' и 'zbmath', а также переопределен smarty-шаблон пользовательского интерфейса, чтобы позволить пользователям с правами менеджера журнала, а также самим рецензентам добавлять или изменять эти поля.

ORCID ID	<input type="text" value="http://orcid.org/0000-0001-7789-23"/>																					
MathSciNet	<input type="text" value="https://mathscinet.ams.org/mathscin"/>																					
ZBMath	<input type="text" value="https://zbmath.org/?q=al%3Allipaché"/>																					
URL	<input type="text"/>																					
Phone	<input type="text"/>																					
Fax	<input type="text"/>																					
Reviewing interests	<table border="1"> <tr> <td>00A30 x</td> <td>31B15 x</td> <td>35A01 x</td> <td>35A02 x</td> <td>35A35 x</td> <td>35C05 x</td> <td>35J05 x</td> </tr> <tr> <td>35J25 x</td> <td>35P25 x</td> <td>35Q60 x</td> <td>41A15 x</td> <td>45B05 x</td> <td>45F05 x</td> <td>45G05 x</td> </tr> <tr> <td>65R20 x</td> <td>65Z05 x</td> <td>68T30 x</td> <td>78A45 x</td> <td>94A08 x</td> <td>94A14 x</td> <td></td> </tr> </table>	00A30 x	31B15 x	35A01 x	35A02 x	35A35 x	35C05 x	35J05 x	35J25 x	35P25 x	35Q60 x	41A15 x	45B05 x	45F05 x	45G05 x	65R20 x	65Z05 x	68T30 x	78A45 x	94A08 x	94A14 x	
00A30 x	31B15 x	35A01 x	35A02 x	35A35 x	35C05 x	35J05 x																
35J25 x	35P25 x	35Q60 x	41A15 x	45B05 x	45F05 x	45G05 x																
65R20 x	65Z05 x	68T30 x	78A45 x	94A08 x	94A14 x																	
Private Notes	<input type="text"/>																					
Mailing Address	<input type="text"/>																					

Рис. 4 — Измененный пользовательский интерфейс менеджера журнала

Далее коды MSC, указанные в этих статьях (см. рис. 7), выделяются используя технологию XPath и технику регулярных выражений и включаются в профиль эксперта. Кроме того, если требуется, выполняется преобразование кодов классификаторов MSC2000 в коды MSC2010.

Как правило, основным кодом, определяющим тематику статьи, является первый код, также могут быть указаны вторичные коды. Таким образом, также можно задать первичному классификатору статьи более высокий вес (например, вдвое больше вторичного). Тогда получим следующую формулу расчета веса MSC кода:

$$w(msc_i, Reviewer_j) = \sum (g * tf_{primary}(msc_i, Reviewer_j) + (1 - g) * tf_{primary}(msc_i, Reviewer_j))$$

Вычислив вес каждого извлеченного кода классификатора относительно экспертов по формуле выше, запишем его в соответствующий им профиль. Далее, из статей, в процессе рецензирования которых участвовал эксперт (как рецензент, составитель тематического выпуска или член редколлегии, представивший статью), извлекаются коды MSC, выполняется преобразование в коды MSC2010, после чего эти коды добавляются в профиль рассматриваемого эксперта. Если код классификатора уже присутствует в профиле, увеличивается вес этого кода. В случае, если эксперт, просмотрев статью, выбрал опцию “UNABLE TO DO THE REVIEW” (отказ от рецензирования) на странице “Review” информационной системы OJS, то у кодов в профиле эксперта, совпадающих с кодами статьи, вес уменьшается.

Очевидно, что эффективность работы подхода определяется покрытием экспертами всей области тематики журнала. В то же время, MSC классификаторы имеют трехуровневую, вложенную структуру. Таким образом, можно рассматривать соответствующие экспертам коды как три разных, и в зависимости от ее узконаправленности можно каждому классу задать свой весовой коэффициент. Тогда даже в малопокрытых областях система сработает.

В соответствии с методом tf-idf, следующим шагом необходимо вычислить tf-idf для собранных классификаторов по трем группам: первый слой, использующий первые два символа кода, второй использующий три и третий использующий классификатор целиком.

В работе использована следующая формула расчета tf-idf:

$$tf(t, d) \times idf(t, D) = \left( 0.5 + \left( 0.5 + \frac{f_{t,d}}{\max_{t' \in d} f_{t' \in d}} \right) \right) \times \log \frac{N}{df}$$

где  $N$  – общее число экспертов в системе,  $df$  – число экспертов из  $D$ , у которых в списке интересов есть хотя бы один код, а в качестве  $f_{t,d}$  будем использовать вес, полученный на предыдущем шаге. Это значение нормируется по максимуму для каждого эксперта  $d$ , используется он для сглаживания разницы между экспертами, у которых мало научных интересов и тех, у кого их много.

Теперь, вычислив значение tf-idf для каждой пары MSC-код – эксперт, получаем векторное представление метода. Остается вычислить косинусную меру между вектором запроса, с каждым экспертом в системе.

Вычислив косинусную меру для каждого слоя классификатора, просуммируем, используя уменьшающийся коэффициент для каждого слоя

$$score(msc_{ik}, Reviewer_j) = similarity \times \delta,$$

где  $\delta$  – весовой коэффициент, равный 0.5 если полное соответствие, 0.3 – совпали первые три символа, и 0.2 – при совпадении первых двух символов. Полученный результат сортируется по убыванию и отображается в виде процентного соотношения. Список рекомендаций, сформированный в автоматическом режиме, представлен на рис. 5.

Назначение рецензентов в системе OJS производится пользователем, имеющим роли «Редактор» или «Редактор раздела». Выбор рецензентов производится из списка пользователей системы, имеющих роль «Рецензент». Также в списке выводится описание двух наиболее часто встречающихся кодов MSC.

Разработанный метод автоматизированного подбора возможных рецензентов для научных работ, поступающих для публикации в информационную систему научного журнала, основан на использовании Mathematics Subject Classification 2010 для определения предметной области, к которой относятся результаты исследования, представленные к публикации. Метод реализован в виде сервиса по технологии плагинов OJS и может быть использован в любом математическом журнале, функционирующем на платформе OJS, это существенно упрощает работу редакторскую работу в информационной среде. Метод апробирован в информационной системе научного журнала Lobachevskii Journal of Mathematics.

The screenshot displays the OJS submission review interface for submission #843. The main section is titled "Peer Review Round 1" and includes a "Select Reviewer" tool. The tool allows filtering reviewers by "Reviewing interests" (set to "contains") and a search box. Below the search box, a list of reviewers is shown with their names, reviewing interests, recommendation percentages, and scores. The reviewers are listed in a table format.

NAME	REVIEWING INTERESTS	RECOMMENDATION	DONE
<a href="#">[REDACTED]</a>	General topology (Spaces with richer structures); Differential geometry (Global differential geometry)	67.0%	8
<a href="#">[REDACTED]</a>	Global analysis, analysis on manifolds (Pseudogroups, differentiable groupoids and general structures on manifolds); Partial differential equations (General topics)	36.6%	1
<a href="#">[REDACTED]</a>	Global analysis, analysis on manifolds (General theory of differentiable manifolds); Differential geometry (Global differential geometry)	35.2%	7
<a href="#">[REDACTED]</a>	Differential geometry (Global differential geometry); Global analysis, analysis on manifolds (General theory of differentiable manifolds)	33.5%	7
<a href="#">[REDACTED]</a>	Functions of a complex variable (Miscellaneous topics of analysis in the complex domain); Partial differential equations (Elliptic equations and systems)	29.2%	0

Рис. 5 — Рекомендации по подбору рецензентов для выбранной статьи

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Выполнен анализ исследуемой области интеллектуальной обработки математических знаний в издательских информационных системах.
2. Сформирован комплекс информационных систем, который обеспечивает реализацию различных сервисов управления цифровыми коллекциями, научными изданиями и информационную поддержку научно-исследовательской деятельности в области математики.
3. Разработаны программные средства автоматизированной трансформации цифровых коллекций математических научных документов в цифровые форматы, допускающие применение методов текстовой аналитики, а именно конвертер документов в формате OpenXML в T<sub>E</sub>X.
4. Реализован программный комплекс автоматизированной семантической обработки физико-математических документов (например, электронных версий сборников трудов), с возможностью ее применения на большие коллекции.
5. Исследованы методы интеграции выделенных в п.4. метаданных в онтологии, описывающие структуру документов (например, DoCo).
6. Разработан программный модуль интеграции выпусков журналов, работающих на издательской информационной системе Open Journal Systems, в международную базу цитирования РИНЦ.
7. Разработана рекомендательная система автоматизированного подбора индексов Универсальной десятичной классификации (УДК) для физико-математических документов.
8. Разработана рекомендательная система подбора рецензентов для научных документов в информационных издательских системах.

Разработанный комплекс программных средств может существенно упростить работу редакции научных журналов, за счет увеличения автоматизированных процессов. Несомненная положительная особенность системы определяется тем, что программные модули не являются узкоспециализированными. Их применение возможно в различных областях деятельности, где требуется выделение семантической оснащенности текста на естественном языке и преобразование его в требуемый формат. Все описанные методы были использованы автором и получили свою программную реализацию при работе в редакции журналов «Lobachevskii Journal of Mathematics» и «Электронные библиотеки» (Казанский (Приволжский) федеральный университет).

## Публикации автора по теме диссертации

### В изданиях, входящих в международные базы цитирования WoS и Scopus

1. *Elizarov, A. M.* Automated system of services for processing of large collections of scientific documents / A. M. Elizarov, S. M. Khaydarov, E. K. Lipachev // 18th International Conference on Data Analytics and Management in Data Intensive Domains, (DAMDID/RCDL 2016). Vol. 1752. — CEUR Workshop Proceedings, 2016. — P. 58–64. — URL: <http://ceur-ws.org/Vol-1752/paper10.pdf>. — (авт. вклад, 0,4375 п. л.)
2. *Elizarov, A. M.* Scientific documents ontologies for semantic representation of digital libraries / A. M. Elizarov, S. M. Khaydarov, E. K. Lipachev // Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017). — IEEE, 2017. — P. 1–5. — DOI: 10.1109/RPC.2017.8168064. — URL: <http://ieeexplore.ieee.org/document/8168064/>. — (авт. вклад, 0,3125 п. л.)
3. Services for formation of digital documents metadata in the formats of international science-based databases / A. M. Elizarov, S. M. Khaydarov, E. K. Lipachev, N. V. Zaitseva, D. S. Zuev // 20th Conference Scientific Services and Internet (SSI 2018). Vol. 2260. — CEUR Workshop Proceedings, 2018. — P. 175–185. — URL: [http://ceur-ws.org/Vol-2260/53\\_175-185.pdf](http://ceur-ws.org/Vol-2260/53_175-185.pdf). — (авт. вклад, 0,6875 п. л.)
4. *Khaydarov, S. M.* Recommender system of physical and mathematical documents classification / S. M. Khaydarov, G. S. Yamalutdinova // 20th Conference Scientific Services and Internet (SSI 2018). T. 2260. — CEUR Workshop Proceedings, 2018. — С. 480–486. — URL: [http://ceur-ws.org/Vol-2260/57\\_480-486.pdf](http://ceur-ws.org/Vol-2260/57_480-486.pdf). — (авт. вклад, 0,4375 п. л.)
5. *Elizarov, A. M.* The formation method of recommendations in the process of scientific peer review of mathematical papers / A. M. Elizarov, S. M. Khaydarov, E. Lipachev // 21st Conference on Scientific Services and Internet (SSI 2019). T. 2260. — CEUR Workshop Proceedings, 2020. — С. 126–135. — URL: <http://ceur-ws.org/Vol-2543/rpaper12.pdf>. — (авт. вклад, 0,625 п. л.)

### Свидетельства о регистрации программ ЭВМ, приравняемые ВАК к публикации в рецензируемых изданиях

6. *Елизаров, А. М.* Программный комплекс выделения метаданных из коллекций физико-математических документов, представленных в формате OpenXML: свидетельство о государственной регистрации программы для ЭВМ №2016618802 Российская Федерация, 08.08.16 / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2016616511 от 21.06.16. — (авт. вклад, 0,1 п. л.)

7. *Елизаров, А. М.* Программный комплекс автоматизированного формирования оригинал-макетов научных изданий из документов в форматах .dosh, .doc, .rtf, .odt: свидетельство о государственной регистрации программы для ЭВМ №2017610506 Российская Федерация, 12.01.17 / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2016662589 от 18.11.16. — (авт. вклад, 0,1 п. л.)
8. Программный комплекс формирования рекомендаций по подбору рецензентов для научных документов в информационных издательских системах: свидетельство о государственной регистрации программы для ЭВМ №2018611617 Российская Федерация, 02.02.2018 / Д. Ю. Ахметов, А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2017662838 от 11.12.2017. — (авт. вклад, 0,1 п. л.)
9. *Елизаров, А. М.* Программа автоматизированного формирования метаданных в формате Российского индекса научного цитирования для статей журнала «Электронные библиотеки»: свидетельство о государственной регистрации программы для ЭВМ №2018612458 Российская Федерация, 16.02.2018 / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2017663206 от 19.12.2017. — (авт. вклад, 0,1 п. л.)
10. Программа автоматизированного формирования метаданных документов цифровой математической библиотеки Lobachevskii DML: свидетельство о государственной регистрации программы для ЭВМ №2019611328 Российская Федерация, 24.01.2019 / А. М. Елизаров, Н. В. Зайцева, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2019610406 от 15.01.2019. — (авт. вклад, 0,1 п. л.)
11. *Елизаров, А. М.* Программа автоматизированного формирования выпусков журнала «Электронные библиотеки»: свидетельство о государственной регистрации программы для ЭВМ №2020610082 Российская Федерация, 09.01.2020 / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров. — заявитель и правообладатель ФГАОУ ВО КФУ; заявка №2019666967 от 20.12.2019. — (авт. вклад, 0,1 п. л.)

### **В сборниках трудов конференций**

12. *Липачёв, Е. К.* Система сервисов преобразования электронных математических документов на основе облачных технологий / Е. К. Липачёв, Ш. М. Хайдаров // Тр. Матем. центра им. Н.И. Лобачевского. Лобачевские чтения – 2013: материалы 12-й молодежной школы-конференции (Казань, 24-29 октября 2013 г.) Т. 47. — Казань : Изд-во Казан. матем. об-ва, 2013. — С. 139–140. — (авт. вклад, 0,125 п. л.)

13. *Хайдаров, Ш. М.* Методы управления математическим контентом в информационных издательских системах / Ш. М. Хайдаров // Тр. Матем. центра им. Н.И. Лобачевского. Лобачевские чтения – 2015: материалы 14-й молодежной научной школы-конференции. Т. 52. — Казань : Изд-во Казан. матем. об-ва, 2015. — С. 162–165. — (авт. вклад, 0,25 п. л.)
14. *Елизаров, А. М.* Семантический анализ больших коллекций научных документов / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Труды международной конференции по компьютерной и когнитивной лингвистике TEL–2016. — Казань : Изд-во Казан. ун– Та, 2016. — С. 21–25. — (авт. вклад, 0,3125 п. л.)
15. Методы автоматизированного извлечения метаданных научных публикаций для библиографических и реферативных баз цитирования / А. Н. Герасимов, А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Информационное общество: образование, наука, культура и технологии будущего: сборн. научных статей. Тр. XIX Объед. конф. «Интернет и современное общество» (IMS–2016), СПб, 22–24 июня 2016 г. — СПб : Ун-т ИТМО, 2016. — С. 41–48. — (авт. вклад, 0,5 п. л.)
16. *Хайдаров, Ш. М.* Методы формирования метаданных научных коллекций / Ш. М. Хайдаров // Тр. Матем. центра им. Н.И. Лобачевского. Лобачевские чтения – 2016: материалы 15-й молодежной научной школы-конференции. Т. 53. — Казань : Изд-во Казан. матем. об-ва, Изд-во Академии наук РТ, 2016. — С. 161–162. — (авт. вклад, 0,125 п. л.)
17. *Хайдаров, Ш. М.* Методы извлечения метаданных научных публикаций на основе структурного и семантического анализа документов / Ш. М. Хайдаров // Материалы XVII Всероссийской конф. молодых учёных по мат моделированию. г. Новосибирск, Россия, 30 октября – 3 ноября 2016 г. — Новосибирск : ИВТ СО РАН, 2016. — С. 110–111. — (авт. вклад, 0,125 п. л.)
18. *Хайдаров, Ш. М.* Метод автоматического описания структуры документов математической коллекции на основе онтологий / Ш. М. Хайдаров // Тр. Матем. центра им. Н.И. Лобачевского. Лобачевские чтения – 2017: материалы 16-й молодежной научной школы-конференции. Т. 55. — Казань : Изд-во Казан. ун-та, 2017. — С. 150–153. — (авт. вклад, 0,25 п. л.)
19. Методы интеграции цифрового физико-математического контента в электронных библиотеках / А. М. Елизаров, Д. С. Зуев, Е. К. Липачёв, Ш. М. Хайдаров // Международный научный семинар «Нелинейные модели в механике, статистике, теории поля и космологии» – GRACOS-17. Международная школа по математическому моделированию в системах компьютерной математики – «KAZCAS-2017». Международная научно-практическая конференция – «ИТОН-2017»: материалы семинара, школы и конференции. — Казань : Изд-во Академии наук РТ, 2017. — С. 166–170. — (авт. вклад, 0,3125 п. л.)

20. *Хайдаров, Ш. М.* Система автоматизированного формирования выпусков научных журналов / Ш. М. Хайдаров // Труды Матем. центра им. Н.И. Лобачевского. Лобачевские чтения – 2019: материалы XVIII Всероссийской молодежной научной школы-конференции. Т. 58. — Казань : Изд-во Академии наук РТ, 2019. — С. 201–203. — (авт. вклад, 0,1875 п. л.)
21. *Зайцева, Н. В.* Плагин автоматизированного формирования метаданных документов цифровой математической библиотеки Lobachevskii DML / Н. В. Зайцева, Ш. М. Хайдаров // Международная школа «Математическое моделирование фундаментальных объектов и явлений в системах компьютерной математики» – «KAZCAS–2018»: лекции и материалы школы. — Казань : Изд-во Академии наук РТ, 2018. — С. 129. — (авт. вклад, 0,0625 п. л.)

### **В прочих научных изданиях**

22. *Хайдаров, Ш. М.* Семантический анализ документов в системе управления цифровыми научными коллекциями / Ш. М. Хайдаров // Электронные библиотеки. — 2015. — Т. 18, № 1/2. — С. 61–85. — URL: <https://elbib.kpfu.ru/ru/article/369>. — (авт. вклад, 1,5625 п. л.)
23. *Елизаров, А. М.* Автоматизированная система структурной и семантической обработки физико-математического контента / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Ученые записки ИСГЗ. — 2016. — Т. 1, № 14. — С. 210–215. — (авт. вклад, 0,375 п. л.)
24. *Елизаров, А. М.* Структура и сервисы цифровой математической библиотеки Lobachevskii-DML / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Ученые записки ИСГЗ. — 2017. — Т. 1, № 15. — С. 215–220. — (авт. вклад, 0,375 п. л.)
25. *Хайдаров, Ш. М.* Алгоритм формирования словарей рекомендуемой системы подбора классификаторов научной информации / Ш. М. Хайдаров, Г. Ш. Ямалутдинова // Ученые записки ИСГЗ. — 2017. — 1 (15). — С. 552–557. — (авт. вклад, 0,375 п. л.)
26. Open Journal Systems в практике работы гуманитарного научного журнала / М. С. Галявиева, А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Вестник Казанского государственного университета культуры и искусств. — 2017. — № 1. — С. 30–34. — (авт. вклад, 0,3125 п. л.)
27. *Елизаров, А. М.* Рекомендательная система поиска экспертов для проведения научного рецензирования в математическом журнале / А. М. Елизаров, Е. К. Липачёв, Ш. М. Хайдаров // Электронные библиотеки. — 2020. — Т. 23, № 4. — С. 708–732. — DOI: 10.26907/1562-5419-2020-23-4-708-732. — (авт. вклад, 1,5625 п. л.)