

УДК 004.4  
ББК 30ф

**Елизаров А. М., Липачёв Е. К., Хасьянов А.Ф.**

Институт математики и механики им. Н.И. Лобачевского,  
Высшая школа информационных технологий и информационных систем  
Казанского (Приволжского) федерального университета  
amelizarov@gmail.com, elipachev@gmail.com, ak@it.kfu.ru

## **МОДЕЛЬ ЦИФРОВОЙ ИНФРАСТРУКТУРЫ НАУЧНО- ОБРАЗОВАТЕЛЬНОГО КЛАСТЕРА РЕСПУБЛИКИ ТАТАРСТАН**

*Аннотация:* Предложена модель цифровой инфраструктуры научно-образовательного кластера Республики Татарстан (РТ) как элемента цифровой экономики РТ, активно формируемой в настоящее время. Создание такой инфраструктуры нацелено на увеличение роли электронных научных библиотек путем использования семантических технологий, в том числе для анализа больших неоднородных данных на основе методов текстового анализа и тематического моделирования. Представлены методы семантического управления цифровыми коллекциями научных материалов. Проектирование модели выполнено в соответствии с основными известными и хорошо апробированными принципами создания глобальных электронных научных библиотек.

*Ключевые слова:* электронные библиотеки, извлечение метаданных, семантическое структурирование контента, семантические связи информационных объектов, научное информационное пространство.

**Elizarov A. M., Lipachev E. K., Khasianov A.F.**

N. I. Lobachevskii Institute of Mathematics and Mechanics,  
Higher School of Information Technologies and Information Systems  
Kazan (Volga Region) Federal University  
Kazan, Russia  
amelizarov@gmail.com, elipachev@gmail.com, ak@it.kfu.ru

## **MODEL OF DIGITAL INFRASTRUCTURE OF THE SCIENTIFIC AND EDUCATIONAL CLUSTER OF THE REPUBLIC OF TATARSTAN**

*Abstract:* We represent the model of the digital infrastructure of the scientific and educational cluster of the Republic of Tatarstan. This infrastructure is an element of the digital economy of the Republic of Tatarstan, which is currently being formed. The creation of such an infrastructure is aimed at increasing the role of electronic scientific libraries by using semantic technologies, including for analyzing large heterogeneous data based on text analysis and modeling methods. We present methods of semantic management of digital collections of scientific mate-

rials. The design of the model is carried out in accordance with the basic principles of the creation of global electronic scientific libraries.

**Keywords:** *digital libraries, extraction of metadata, semantic structuring of content, semantic links of information objects, scientific information space.*

**Введение.** Фундаментальная научная проблема, разработке которой посвящена настоящая работа, связана с созданием семантических методов консолидации и управления электронными коллекциями научно-образовательных материалов из различных предметных областей, а также их интеграцией в мировое научно-образовательное пространство. Нами спроектирована модель цифровой инфраструктуры научно-образовательного кластера РТ, которая разработана в соответствии с основными принципами, заложенными в ведущих международных проектах создания глобальных цифровых библиотек (Digital Humanities, Digital Mathematical Library и др.). [1, 2]. К числу основных решаемых задач относятся: формирование современной цифровой инфраструктуры на всех уровнях проведения научных исследований, предоставление на ее основе всевозможных информационных сервисов и обеспечение пользователям персонифицированного доступа к информации, а также расширение спектра научных коммуникаций.

Традиционные библиотеки связывают документы на основе библиографических ссылок, а поиск документа осуществляется на основе его библиографического описания и выбранных тематических рубрик, что не обеспечивает автоматизации поисковых процессов и непосредственного доступа пользователя к элементам знания. Появление электронных (цифровых) библиотек обеспечило новый уровень работы с документами – созданы информационные системы, реализующие сервисы поиска и обработки документов. Основными элементами создаваемой цифровой инфраструктуры будут уже не документы, а информационные объекты, а также логические связи между ними. Важна также интеграция отечественных информационных ресурсов в мировые базы данных, для этого необходим целый ряд специализированных сервисов (см., например, [3]). К таковым, в частности, относятся сервисы сбора наукометрических данных, учет которых в последнее время стал повсеместной практикой в научных и образовательных организациях.

В настоящее время создание электронных научных библиотек актуально для развития научной цифровой инфраструктуры любого региона. Республика Татарстан – один из российских регионов с развитой научно-исследовательской инфраструктурой, поэтому решение обозначенных проблем актуально для РТ.

Прототипом создаваемой цифровой инфраструктуры служит информационная система [science.tatarstan.ru](http://science.tatarstan.ru), которая развивается нами, начиная с 2013 года, и предоставляет возможности размещения, хранения и обработки научной информации гуманитарного и естественнонаучного содержания (см. [4]). Следующий этап в развитии этой системы – разработка и использование методов семантического управления цифровыми коллекциями научных материалов различной тематики.

**Анализ современного состояния исследований в данной области.** Такой анализ показал, что в настоящее время научное сообщество, становясь все более глобальным, заинтересовано в максимально быстром распространении полученного нового знания и оперативном доступе к нему, что способны обеспечить только новые формы научной коммуникации (см., например, [5]). Идеи проведения научных исследований в распределенной цифровой среде отражают сегодня такие новые понятия, как «eScience», «eResearch», «eHumanities» и др. (см. [1]).

Перевод знаний в цифровую форму и перемещение научных коммуникаций в сетевое пространство не только изменили сложившуюся научную инфраструктуру, но и поставили новые задачи по управлению знаниями. Как известно, необходимым элементом научного исследования является описание связей новых научных результатов с ранее полученными. В современных условиях для выполнения этого требования необходимо наличие в сети научного контента, как современного, так и ставшего уже классическим.

Частью формируемой новой научной цифровой инфраструктуры являются электронные библиотеки. Примерами таких библиотек являются Общероссийский математический портал Math-Net.Ru (<http://www.mathnet.ru/>), Czech Digital Mathematics Library (DML-CZ, <http://dml.cz/>), Centre de Diffusion de Revues Académiques Mathématiques (CEDRAM, <http://www.cedram.org/>), Numerisation de Documents Anciens Mathématiques (NUMDAM, <http://www.cedram.org/>). По этим же принципам нами разрабатывается электронная библиотека Lobachevskii Digital Mathematical Library (Lobachevskii DML, <http://www.lobachevskii-dml.ru/>, см. также [6]). В [7] проведен анализ существующих электронных математических библиотек с точки зрения модели DELOS Digital Library Reference Model. Европейский проект EuDML (<https://eudml.org/>) направлен на интеграцию европейских математических ресурсов, а инициатива World Digital Mathematical Library ставит основной задачей объединение в распределенной системе всех цифровых коллекций электронных математических документов [2].

Современные электронные научные библиотеки содержат наборы документов с различными структурой и форматами данных, что затрудняет организацию информационных сервисов для семантической обработки информации. Разработка в электронной библиотеке таких сервисов приводит к образованию многослойной семантической структуры контента [8]. Такая структура может формироваться на основе онтологии связей и служит источником информации для проведения качественно новых наукометрических измерений.

**Цель и основные задачи проводимого исследования.** Основная цель – создание методов семантического управления научными цифровыми коллекциями и реализация на их основе прототипа цифровой научной библиотеки, спроектированной в соответствии с основными принципами, разработанными в ведущих проектах создания глобальных цифровых научных библиотек. В настоящее время в РФ отсутствует цифровая библиотека, способная консолидировать управление такими цифровыми ресурсами.

Основные задачи, решаемые нами:

1) Приведение электронных коллекций научных документов к единому цифровому формату, допускающему применение методов текстовой аналитики, в частности, выделение метаданных. Для этого используются методы автоматической обработки доступных электронных версий сборников трудов научных конференций, проведенных в РТ, оцифрованных (электронных) архивов издаваемых научных журналов, отчёты по научным темам АН РТ.

2) Создание модели цифровой научной библиотеки, объединяющей электронные ресурсы, указанные выше, и основанной на семантических связях между информационными объектами электронных документов. Для описания семантики связей информационных объектов создаются соответствующие словари и используются существующие онтологии связей (SPAR, SKOS, CERIF, DoCo и др.) (см. [9]).

**Предлагаемые подходы и методы.** Создание и распространение научных знаний и историко-культурной информации – стратегически важные направления инновационного развития любого региона, в том числе Республики Татарстан. Для полноценного включения в мировые информационные потоки необходимо решить задачу интеграции в едином информационном пространстве электронных документов научного и образовательного содержания. Один из подходов к решению этой задачи основан на создании специализированных информационных систем. Интеграция электронных документов дает возможность обращения к данным независимо от места их размещения; анализа информации, включающего определение характеристик, взаимосвязей и источников; преобразования информации для ее обогащения и приведения в соответствие заданным целям, а также объединения информации, чтобы сделать ее доступной людям, процессам и приложениям. Реализация нашего подхода основана на программной платформе [science.tatarstan.ru](http://science.tatarstan.ru), которая построена в виде веб-портала и реализует документооборот в научных журналах гуманитарного и естественнонаучного циклов. Новизна примененных подходов заключается в рассмотрении цифровых коллекций в соответствии с объектной моделью управления научными данными, формирование которой только началось в ряде европейских проектов. Интеграция цифровых научных коллекций Республики Татарстан базируется на организации семантических связей, методах управления научным контентом и обработки математических коллекций, уже апробированных нами при построении систем управления электронными журналами [10]. При автоматической обработке больших коллекций научных документов применены онтологии описания документов (см. [11]).

**Заключение.** В рамках проводимого исследования создана информационная система облачных сервисов, обеспечивающих многоуровневую поддержку процессов подготовки, публикации и интеграции научных знаний, образовательных и историко-культурных материалов в информационное пространство РТ. Среди этих сервисов: автоматизированное приведение документов цифровых коллекций к единому формату; оптимизация автоматической обработки коллекций с помощью разработанных шаблонов представ-

ления электронных научных документов; автоматизированная тематическая кластеризация научных материалов цифровых коллекций; автоматизированная экстракция метаданных из документов цифровых коллекций электронной библиотеки и формирование наборов метаописаний в форматах международных наукометрических баз данных.

Основные полученные и ожидаемые результаты проводимых исследований внесут вклад в развитие современных моделей хранения и распространения научных знаний. Как естественное развитие и расширение платформы science.tatarstan.ru, эти результаты позволят увеличить количество электронных научных коллекций, содержащих документы различной тематической направленности, которые станут составной частью цифровой научной библиотеки РТ.

**Благодарность.** Работа выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения государственного задания в сфере научной деятельности, проект 1.2368.2017/ПЧ, и при частичной финансовой поддержке РФФИ и Правительства Республики Татарстан в рамках научного проекта № 18-47-160012.

### **Литература**

- [1] Thaller M. Controversies around the Digital Humanities: An Agenda // Historical Social Research /Historische Sozialforschung. 2012. V. 37, No 3. P. 7–15.
- [2] Developing a 21st Century Global Library for Mathematics Research. Washington, The National Academies Press, 2014. 131 p. DOI: <https://doi.org/10.17226/18619>.
- [3] Елизаров А.М., Зуев Д.С., Липачёв Е.К. Управление жизненным циклом электронных публикаций в информационной системе научного журнала // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2014. № 4. С. 81–88.
- [4] Ахметов Д.Ю., Герасимов А.Н., Грачев А.О., Елизаров А.М., Липачёв Е.К. Облачная платформа поддержки электронных научных изданий // Ученые записки ИСГЗ. 2014. № 1-1 (12). С. 13–19.
- [5] Щур Л.Н. Роль инфокоммуникационных технологий в развитии процесса глобализации научных исследований // Информационное общество. 2012. № 5. С. 16–24.
- [6] Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М. Структура и сервисы цифровой математической библиотеки Lobachevskii-dml // Ученые записки ИСГЗ. 2017. № 1 (15). С. 215–220.
- [7] Elizarov A.M., Lipachev E.K., Zuev D.S. Digital Mathematical Libraries: Overview of Implementations and Content Management Services // Data Analytics and Management in Data Intensive Domains: Collection of Scientific Papers of the XIX International Conference DAMDID / RCDL'2017. Moscow: FRC CSC RAS, 2017. P. 394–402. CEUR Workshop Proceedings, 2017. Vol. 2022. P. 317–325.

- [8] Kogalovsky M.R., Parinov S.I. Scholarly Communication in a Semantically Enrichable Research Information System with Embedded Taxonomy of Scientific Relationships. In: Klinov P., Mouromtsev D. (eds) Knowledge Engineering and Semantic Web. Communications in Computer and Information Science, 2015, V. 518. Springer, Cham. doi: 10.1007/978-3-319-24543-0\_7.
- [9] Peroni S. Semantic Web Technologies and Legal Scholarly Publishing. Volume 15 of the Law, Governance and Technology Series. Cham, Switzerland: Springer, 2014. doi: 10.1007/978-3-319-04777-5.
- [10] Галявиева М.С., Елизаров А.М., Липачёв Е.К. Цифровая инфраструктура электронного научного журнала: автоматизация редакционно-издательских процессов и система сервисов // Электронные библиотеки. 2016. Т. 19. № 5. С. 408–465.
- [11] Elizarov A.M., Khaydarov Sh.M., Lipachev E.K. Scientific documents ontologies for semantic representation of digital libraries // Second Russia and Pacific Conference on Computer Technology and Applications (RPC). Vladivostok, Russky Island, Russia 25–29 September. IEEE, 2017. P. 1–5. doi: 10.1109/RPC.2017.8168064.