

differential expression with polynomial coefficients in a Hilbert space $\mathcal{L}^2(-\infty, +\infty)$. The resulting matrices are generalized Jacobian matrices. Based on this, we study the problem of deficiency index and the spectral characteristics of self-adjoint extensions of the minimal operator generated by differential expressions with polynomial coefficients, when the matrices, associated with them, are purely diagonal or only elements of the two diagonals are non-zero.

Keywords: symmetric differential and difference operators, generalized Jacobi matrix, deficiency index, self-adjoint extensions, spectrum.

УДК 004.42

МЕТОД ОРГАНИЗАЦИИ ЦИФРОВЫХ КОЛЛЕКЦИЙ НА ОСНОВЕ МЕТАДААННЫХ И СИСТЕМЫ СЕМАНТИЧЕСКИХ СВЯЗЕЙ

Р.Р. Батыршина¹, Е.К. Липачев², Э.М. Сабитова³, К.А. Семенова⁴

¹ r-batyrshina@mail.ru; Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

² elipachev@gmail.ru; Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

³ www.insli1234@mail.ru; Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

⁴ [kseniia_andreevna95@mail.ru](mailto:kсениа_andreевна95@mail.ru); Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

Работа посвящена методам организации цифровых математических коллекций из набора документов, хранящихся в различных файловых форматах, имеющих разную структуру (напр., монографии, сборники статей, тезисы докладов) и отличающихся по составу представленных метаданных. Описан алгоритм создания цифровой коллекции “Трудов математического центра им. Н.И.Лобачевского” и включения этой коллекции в цифровую математическую библиотеку Lobachevskii-DML. В связи с этим приведены методы извлечения метаданных из документов, установления семантических связей и разделения документов, представляющих сборники, на отдельные статьи.

Ключевые слова: цифровая коллекция, цифровая математическая библиотека, метаданные, семантическая связь, семантический метод, Lobachevskii-DML

Работа посвящена методам создания цифровых научных коллекций из массива разнородных оцифрованных документов в соответствии с основополагающими принципами Всемирной цифровой математической библиотеки [2]. Описан процесс формирования цифровой коллекции из набора файлов, содержащих тома “Трудов математического центра им. Н.И.Лобачевского” за 1998–2017 гг.

Как известно, управление информацией в сети основано на использовании метаданных [3]– [4]. Создание цифровой коллекции из набора документов предполагает не только расширение набора метаданных, но и их нормализацию в соответствии с установленными схемами данных. Цифровые библиотеки, в свою очередь, накладывают дополнительные требования на состав и формат метаданных [5]. Специфика математических документов должна учитываться при формировании метаданных цифровых математических библиотек [6]. Нашей задачей является созда-

ние методов, позволяющих программным способом извлекать необходимые метаданные из цифровых математических документов и устанавливать семантические связи между объектами.

Основным назначением “Трудов математического центра им. Н.И.Лобачевского” является публикация материалов математических конференций и, как следствие, большинство томов “Трудов” содержат несколько десятков статей с ограниченным, с современной точки зрения, составом метаданных. С 1998 года, с момента выпуска первого тома, использовалось несколько правил подготовки материалов, что отразилось на оформлении статей и форматах файлов сборников. Необходимым условием создания цифровой коллекции из файлового массива “Трудов” являлось разделение томов на отдельные статьи, выделение метаданных, описывающих каждую статью, генерация дополнительных метаданных, содержащих, в частности, библиографическое описание статьи, ссылку на файл статьи в цифровой коллекции, а также связи с профилями авторов статьи на академических порталах и наукометрических базах (kpfu.ru, MathNet.ru, Scopus и др.).

На первом этапе производилась обработка массива файлов “Трудов” с целью выделения метаданных, описывающих как том в целом, так и статей, входящих в него. В частности, определялись номера страниц всех статей каждого тома. Для этого разработан алгоритм, использующий структурную однородность каждого тома и стиливую однозначность в оформлении статей в нем, для поиска страниц с названиями статей. Это позволило выделить не только названия статей, но и списки авторов, блоки библиографии и другие метаданные (напр., email, ключевые слова), в случае их наличия в тексте. Предварительно была произведена кластеризация, в результате которой тома были разделены на классы по сходству их структуры и оформления. Для каждого класса разработан набор паттернов регулярных выражений, задающих правила поиска информационных блоков. Основой этого алгоритма является подход, предложенный в [7]. Алгоритм реализован в виде программ на C#, позволяющих обрабатывать файлы в форматах \TeX , OpenXML (.docx) и PDF. \TeX -файлы обрабатывались с помощью стандартных функций, реализующих операции с текстовыми строками. Для работы с PDF-файлами использовались функции библиотек PDFLib (<https://www.pdflib.com>) и iTextSharp (<https://www.nuget.org/packages/iTextSharp/>). Для документов, представленных в виде .docx-файлов, производился разбор файла “word/document.xml”, выделенного из .docx-архива в соответствии с форматом Office OpenXML (см., напр., [8]).

Предложен XML-язык описания цифровых математических коллекций, состоящий из набора тегов и XML-схем, основанных на Journal Archiving and Interchange Tag Suite (<https://jats.nlm.nih.gov/1.2d2/>). В нотации этого языка, на основании данных, полученных на этапе обработки массива файлов, проведено описание коллекции “Трудов Математического центра”.

С помощью методов текстового анализа [9] из документов цифровой коллекции выделены термины, из которых образованы наборы ключевых слов для включения в состав метаданных. Алгоритм извлечения терминов является развитием подхода, предложенного в [10].

Следующий этап создания цифровой коллекции состоял из процедур разделения каждого тома “Трудов” на отдельные статьи, присвоения полученным файлам имен в соответствии с правилами коллекции и установления семантических связей. Алгоритм назначения связей в цифровой коллекции основан на методе работы [11]. Процесс выделения статей производился с помощью программы, разработанной на языке Python с использованием функций библиотеки PyPDF2 (<http://pybrary.net/pyPdf/>) [12].

Ряд метаданных, таких как электронная почта авторов, их аффилиация и др. были импортированы и уточнены из профилей авторов на академических сайтах и базах. В этой процедуре были применены семантические связи, установленные в процессе формирования цифровой коллекции.

Система метаданных, подготовленных в процессе работы приведенного алгоритма, позволила сформировать цифровую коллекцию “Трудов математического центра им. Н.И.Лобачевского” и включить ее в состав цифровой библиотеки Lobachevskii-DML (<https://lobachevskii-dml.ru/>) [13], [14].

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. *Developing a 21st Century Global Library for Mathematics Research*. – Washington, The National Academies Press, 2014. – 131 p. DOI: <https://doi.org/10.17226/18619>.
2. Gartner R. *Metadata. Shaping Knowledge from Antiquity to the Semantic Web*. – Springer, 2016. – 114 p.
3. Sicilia M.-A. (Ed.) *Handbook of Metadata, Semantics and Ontologies*. – World Scientific Publishing Co. Pte. Ltd., 2014. – 570 p.
4. Lubas R., Jackson A., Schneider I. *The Metadata Manual*. – Chandos Publishing, 2013. – 216 p.
5. Alemu G., Stevens B. *An Emergent Theory of Digital Library Metadata*. – Elsevier Ltd., 2015. – 122 p.
6. Elizarov A. M., Lipachev E. K., Zuev D. S. *Digital Mathematical Libraries: Overview of Implementations and Content Management Services // CEUR Workshop Proceedings*. – 2017. – Vol. 2022. – P. 317–325.
7. Elizarov A. M., Khaydarov Sh. M., Lipachev E. K. *Scientific documents ontologies for semantic representation of digital libraries // 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC)*. Vladivostok, 2017. – P. 1–5. – DOI: 10.1109/RPC.2017.8168064.
8. *Standard ECMA-376 Office Open XML File Formats*. URL: <http://www.ecma-international.org/publications/standards/Ecma-376.htm>.
9. Ingersoll G. S., Morton T. S., Farris A. L. *Taming Text. How to Find, Organize, and Manipulate It*. – Manning Publications Co., 2013. – 320 p.
10. Батыршина Р. Р. *Метод извлечения терминов в цифровых математических коллекциях // Тр. Матем. центра им. Н. И. Лобачевского*. – Казань: Изд-во Казан. матем. об-ва, 2017. – Т. 55. – С. 24–26.
11. Сабитова Э. М. *Алгоритм извлечения связей в научных цифровых коллекциях // Тр. Матем. центра им. Н. И. Лобачевского*. – Казань: Изд-во Казан. матем. об-ва, 2017. – Т. 55. – С. 123–126.

12. Семенова К. А. *Алгоритм формирования метаданных цифровых научных коллекций на основе семантического анализа документов*. – Выпускная квалификационная работа (бакалаврская работа), Казань, 2017. – 56 с.
13. Elizarov A. M., Lipachev E. K. *Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University* // CEUR Workshop Proceedings. – 2017. – Vol. 2022. – P. 326–333.
14. Елизаров А. М., Липачев Е. К. *Семантические методы и инструменты электронной математической библиотеки Lobachevskii-DML* // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции. – М.: ИПМ им. М. В. Келдыша, 2017. – С. 130–136. – doi:10.20948/abrau-2017-73.

METHOD OF ORGANIZING DIGITAL COLLECTIONS BASED ON METADATA AND SYSTEM OF SEMANTIC RELATIONS

R.R. Batyrshina, E.K. Lipachev, E.M. Sabitova, K.A. Semenova

The work is devoted to the methods of organization of digital mathematical collections from a set of documents stored in various file formats that have different structures (monographs, collections of articles, or abstracts) and different sets of metadata. We describe an algorithm for creating a digital collection of “Proceedings of the N.I. Lobachevskii Mathematical Center” in order to include this collections in the Lobachevskii Digital Mathematical Library. In this regard, we present methods for extracting metadata from documents, establishing semantic relations, as well as methods for splitting documents representing collections into separate articles.

Keywords: digital collection, Digital Mathematics Library, metadata, semantic relation, semantic method, Lobachevskii-DML.

УДК 517.54

ОБРАТНАЯ КРАЕВАЯ ЗАДАЧА АЭРОГИДРОДИНАМИКИ С ЗАДАННОЙ ВЕЛИЧИНОЙ СКОРОСТИ НЕВОЗМУЩЕННОГО ПОТОКА

Р.Б. Салимов¹, Э.Н. Хасанова²

¹ *salimov.rsb@gmail.com*; Казанский государственный архитектурно-строительный университет
² *enkarabasheva@bk.ru*; Казанский государственный архитектурно-строительный университет

Рассмотрена обратная краевая задача аэрогидродинамики в видоизмененной постановке, где требуется найти форму крылового профиля, обтекаемого потенциальным потоком несжимаемой невязкой жидкости, по заданному на нём распределению потенциала скорости как функции абсциссы точки профиля, при заданной величине скорости в передней кромке профиля и величине скорости невозмущенного потока, обтекающего искомый профиль.

Ключевые слова: обратная краевая задача аэрогидродинамики, аналитическая функция, конформное отображение, крыловой профиль.

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 18-31-00060.

Пусть в плоскости комплексного переменного $z = x + iy$ расположен крыловой профиль L_z , обтекаемый потенциальным потоком несжимаемой невязкой жидкости с комплексным потенциалом $w = w(z) = \varphi + i\psi$ и скоростью невозмущенного потока $v_\infty e^{i\eta_\infty}$, $v_\infty > 0$, $\frac{\pi}{2} \leq \eta_\infty \leq \frac{3\pi}{2}$. Примем, что для абсцисс точек L_z , выполняется