

УДК 004.42

## МЕТОД ФОРМИРОВАНИЯ СЕМАНТИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ЦИФРОВЫХ МАТЕМАТИЧЕСКИХ ДОКУМЕНТОВ

П.О. Гафурова<sup>1</sup>

<sup>1</sup> polina\_gafurova@yahoo.com; Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

*Предложен метод семантического представления документов цифровых математической коллекции. Разработаны язык семантического описания документов коллекций, сервис по их обработке, инструменты семантического поиска и представления.*

**Ключевые слова:** семантический веб, семантика, OMDoc, семантическое представление математических документов.

Работа является продолжением исследований, проведенных в [1], и посвящена разработке методов введения семантики в электронные математические документы на основе форматов OMDoc, sTeX и других (см., напр., [2], [3]).

Наличие большого количества источников обязывает представлять средства для обработки информации в сети интернет. Естественно, каждая сфера научного познания имеет особенности. В математике примером таких конструкций являются теоремы, леммы, доказательства, а главное – формулы. Данные конструкции имеют собственные значения и, следовательно, для машинной обработки необходимо включать в них некоторую скрытую информацию, в частности – метаданные. Для обработки математических документов необходимо привести их в семантический вид.

Введение семантики в Web является отличительной чертой Web 3.0. Элементы семантики можно встретить и сейчас, например, в немецкоязычной Wikipedia. Также в качестве примера можно отметить проект KWARC, нацеленный на внедрение семантики в различные документы, и в том числе математические [4], [5]. Для внедрения семантики в математические документы проект KWARC предлагает язык представления математических документов – OMDoc [2], [6]. OMDoc является по своей структуре XML документом с включенными в него математическими формулами. Данный формат был взят нами как основа для разработки языка семантического описания математических документов. Данный формат начал разрабатываться еще в 2000 году и до сих пор полностью не разработан, так как в математических документах существуют проблемы, которые пока решить не удалось [7]. Несмотря на то, что формат начал разрабатываться уже достаточно давно, нет работ, связанных с визуальным представлением документов на этом языке – язык является основой для некоторых других подпроектов проекта KWARC. Однако можно с помощью преобразований осуществлять вывод информации из файла на экран [1], [5].

Для семантического представления формул обычно используется язык MathML, также основанный на XML [8]. MathML работает не во всех браузерах, что затрудняет использование этого языка в качестве языка представления математических документов. Для этого при представлении математических документов

```

<?xml version="1.0" encoding="utf-8"?>
<?xml-stYLESHEET type="text/xsl" href="matth.xsl"?>
<lib>
  <book id="1">
    <meta>
      <author></author>
      <title></title>
      <year></year>
      <publisher></publisher>
    </meta>
    <content>
      <con id="1">
        <contitle></contitle>
        <desk id="1">
          <title></title>
          <sod></sod>
        </desk>
        <ax id="6">
          <title></title>
          <sod></sod>
        </ax>
        <theor id="7">
          <title></title>
          <sod></sod>
          <dok></dok>
          <altdok id="1"></altdok>
          <altdok id="2"></altdok>
          <sled id="1"></sled>
        </theor>
        <ex id="8">
          <title></title>
          <sod></sod>
        </ex>
        <lemm id="9">
          <title></title>
          <sod></sod>
        </lemm>
        <art id="10"></art>
        ...
      </con>
      ...
    </content>
    <literat>
      <bok id="1">
        <num></num>
        <ss></ss>
      </bok>
      ...
    </literat>
  </book>
  ...
</lib>

```

**Рис. 1.** Схема семантического представления документа цифровой коллекции

в веб используют MathJax – язык, который не поддерживает семантику, однако воспроизводится всеми современными браузерами.

Для семантического описания цифровой математической коллекции нами предложен язык, основанный на XML (см., напр., [9]). Этот язык имеет следующую структуру: каждая книга описывается тегом `<book>` и включает в себе: метаданные (теги `<author>`, `<title>`, `<year>`, `<publisher>`), а также тег содержания. В данном теге содержится основная информация по книге без потери структурной составляющей текста – параграфы и главы имеют свое строгое положение в тексте. В параграфах находятся содержательные теги для описания как математических структур: определение `<def>`, лемма `<lemm>`, теорема `<theor>`, аксиома `<ax>`, пример `<ex>`; так и не математических – тег `<article>`. Математические теги также имеют свое строгое местоположение и строение – в каждый математический объект включено краткое название, полный текст или альтернативный текст, также в случае с теоремами приведено доказательство, возможен случай приведения альтернативного доказательства и следствий.

Для передачи математических формул в созданном языке представления используется Presentation MathML (см., напр., [8], [10]). Однако, в связи с проблемами с отображением в браузерах, потребовалось решить задачу преобразования в

MathJax. Для этого было написано приложение на языке Python, которое при чтении из файла кода MathML выполняет преобразование в MathJax (результат на рис. 2). Данное приложение основано на использовании строгой структуры MathML при переформатировании его в MathJax.

```

1 <math>
2 <mfrac>
3 <mrow>
4 <mn>1</mn>
5 <mo>+</mo>
6 <msup>
7 <mo class="MathClass-op">sin</mo>
8 <mn>2</mn>
9 </msup>
10 <mfenced separators="" open="(" close=")">
11 <mi>x</mi><mo>+</mo><mi>y</mi></mfenced>
12 </mrow>
13 <mrow>
14 <mn>2</mn>
15 <mo>+</mo>
16 <mfenced separators="" open="|" close="|">
17 <mrow>
18 <mi>x</mi>
19 <mo>-</mo>
20 <mfrac>
21 <mrow><mn>2</mn><mi>x</mi></mrow>
22 <mrow>
23 <mn>1</mn>
24 <mo>+</mo>
25 <msup><mi>x</mi><mn>2</mn></msup>
26 <msup><mi>y</mi><mn>2</mn></msup>
27 </mrow>
28 </mfrac>
29 </mrow>
30 </mfenced>
31 </mrow>
32 </mfrac>
33 <mo>+</mo>
34 <mi>x</mi>
35 </math>

```

```

1 $\frac{1+\sin^2(x+y)}{2+|x-\frac{2x}{1+x^2}y^2|}+x$

```

Рис. 2. Входные данные и результат работы программы

Следующим шагом является создание веб – приложения на языке PHP с целью вывода документов коллекции на экран и семантического поиска по ним.

Веб-приложение отображает две страницы – первая страница выбора категории и термина (Рис. 3).

### "Семантическая библиотека"

Выберите категорию термина который Вы хотите найти

- ☐ Определение
- ☐ Аксиома
- ☐ Теорема
- ☐ Пример
- ☐ Лемма

Пожалуйста введите термин поиск которого нужно осуществить



Рис. 3. Страница выбора категории и термина

В списке необходимо выбрать в какой категории пользователь хочет совершить поиск, а также вбить в текстовое поле поисковый запрос и нажать на кнопку. Далее происходит поиск по выбранной пользователем категории.

Вторая страница является страницей вывода. При выводе сначала указывается какой термин и из какой категории мы искали. Рис. 4 приведен как результат поиска: выводится название термина, метаданные по документу, такие, как название документа, автор или авторы, год издания. Имеется возможность отображения полного текста, описывающего термин. Данная информация скрыта для пользователя при обычном просмотре. В случае, когда было найдено два или более термина, выводятся все данные термины.

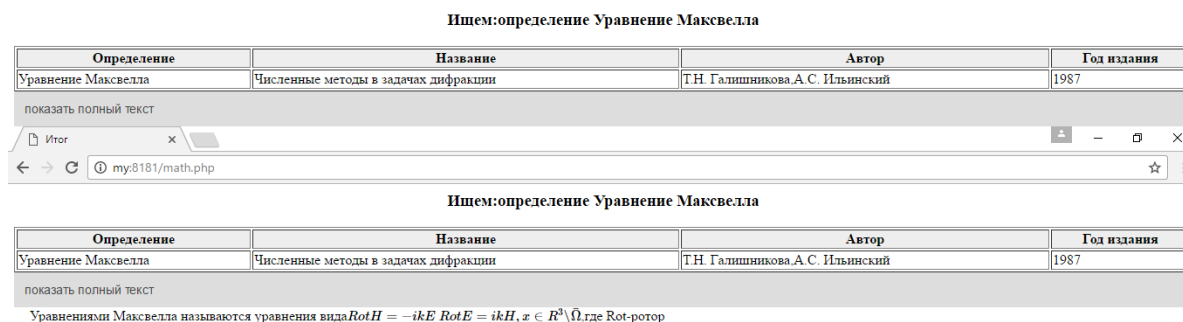


Рис. 4. Страница вывода: случай одного совпадения

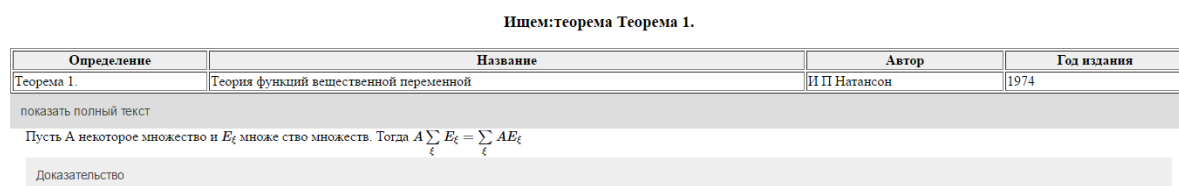


Рис. 5. Страница вывода: случай большего количества совпадений

В случае более сложных конструкций, таких как теоремы или леммы, необходимо также выводить и доказательство. При нажатии на кнопку «показать полный текст» отображается текст теоремы и кнопка, при нажатии на которую будет возможно открытие текста доказательства.

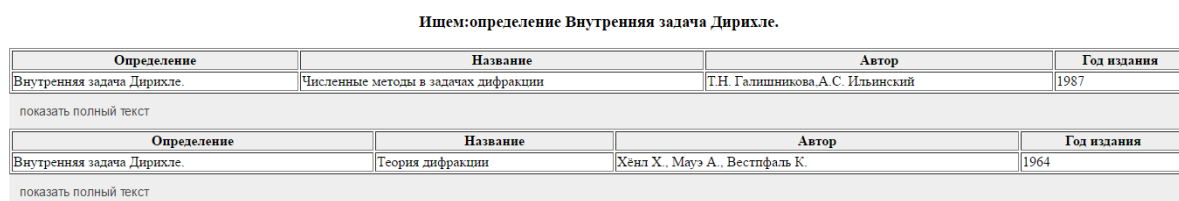


Рис. 6. Страница вывода: случай сложной структуры

В заключение укажем дальнейшие планы по развитию данной работы. Прежде всего, создание средств, позволяющих сохранить семантику при выводе контента, аннотирование терминов и формул.

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 № 15-47-02472.

## Литература

1. Гафурова П.О. Метод обработки математических документов в формате OMDoc // Тр. Матем. центра им. Н. И. Лобачевского. – Казань: Изд-во Казан. матем. об-ва, 2016. – Т. 53. – С. 70–73.

2. Kohlhase M. *OMDoc: An Open Markup Format for Mathematical Documents [version 1.2]* // Lecture Notes in Artificial Intelligence. – 2006. – V. 4180. – 428 p.
3. Kohlhase M. *Using L<sup>A</sup>T<sub>E</sub>X as a Semantic Markup Format*. URL: <https://kwarc.info/kohlhase/papers/mcs08-stex.pdf>.
4. *KWARC: Knowledge Adaptation and Reasoning for Content*. URL: <https://kwarc.info/>
5. Kohlhase M., David C., Ginev D., Cornely J. *eMath 3.0: Building Blocks for a Social and Semantic Web for Online Mathematics & eLearning* // Computer Science, Jacobs University Bremen, Germany. – October 27, 2010. – 14 p.
6. Kohlhase M., Iancu M. *Co-Representing Structure and Meaning of Mathematical Documents* // Computer Science, Jacobs University Bremen, Germany. – October 26, 2015. – 24 p.
7. Kohlhase M. *OMDoc: An Open Markup Format for Mathematical Documents* // FB Informatik, Universitat des Saarlandes D-66041 Saarbrücken, Germany. – 2000. – 64 p.
8. Елизаров А.М., Липачев Е.К., Малахальцев М.А. *Веб-технологии для математика: основы MathML. Практическое руководство* // М.: ФИЗМАТЛИТ, 2010. – 192 с.
9. Елизаров А.М., Липачев Е.К., Малахальцев М.А. *Языки разметки семантического веба. Практические аспекты*. – Казань, 2008.
10. Елизаров А.М., Липачев Е.К., Малахальцев М.А. *Основы MATHML. Представление математических текстов в Internet*. – Казань, 2008.

## METHOD OF SEMANTIC REPRESENTATION DIGITAL MATHEMATIC DOCUMENTS

P.O. Gafurova

*Several works of semantic representation of mathematical documents was reviewed. We also designed the language of semantic representation of mathematical documents, serviced the processing of this language and possibilities for its semantic search and representation of the given language.*

Keywords: semantic Web, semantic, OMDoc, semantic representation of mathematics documents.

УДК 519.7; 512.581

## О МАСКИРОВКЕ АЛГЕБРАИЧЕСКИХ ПЛАТФОРМ В КРИПТОГРАФИИ С ОТКРЫТЫМ КЛЮЧОМ

К.И. Емельянов<sup>1</sup>

<sup>1</sup> [kirillemelyanov11041995@gmail.com](mailto:kirillemelyanov11041995@gmail.com); Казанский (Приволжский) федеральный университет, Институт математики и механики им. Н.И. Лобачевского

*В статье обсуждается маскировка алгебраических платформ, таких как группы и группоиды. Идея подобной маскировка переносится на более сложную платформу – 2-категорию.*

**Ключевые слова:** криптография, маскировка алгебраических платформ, 2-категория.

В работе [1] был предложен метод маскировки алгебраических платформ, позволяющий при некоторых предположениях гарантировать невзламываемость криптографических протоколов в течение любого наперед заданного времени. В работе [2] были описаны некоторые приложения этого метода.