

Обзор инструментов для математических данных.

© А.С. Аджиев

А.М. Атаева

Л.Н. Шиолашвили

Вычислительный центр РАН
lalysh@mathnet.ru

Аннотация

В статье представлен обзор известных инструментов подготовки математических данных, а также различные способы представления в сети: TeX/LaTeX, MathML, различные графические форматы. Рассмотрены основные функции и возможности данных инструментов, их достоинства и недостатки. Дан сравнительный анализ конвертирования математических данных, представленных в форматах TeX/LaTeX и MathML.

Введение

Математическая система обозначений имеет сложную структуру, состоящую из математических формул, выражений, а также текста, к которому может быть применено форматирование различного вида.

Существенным элементом реализации цифровых библиотек научной информации является поддержка математических данных (математических конструкций), с этим связаны вопросы организации создания, хранения, представления математических материалов, обеспечение возможности поиска математических выражений и т.п.

Одним из самых распространенных инструментов создания математических текстов является редакционно-издательская система TeX. Система является разработкой известного математика-программиста Дональда Кнута, представленная в 1986 году как свободно распространяемый продукт. Это была первая система с развитой поддержкой математических формул. Все это и стало причиной представления большинства математических текстов в формате TeX, LaTeX, AMS-TeX.

LaTeX самый распространенный макропакет,

Труды 7^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL '2005, Ярославль, Россия, 2005.

написанный Лесли Лэмпортом и реализующий все необходимые средства для подготовки статей, книг, отчетов и деловых писем.

AMS-TeX пакет, соответствующий стандартам Американского Математического Сообщества.

Основная трудность при внедрении математики в World Wide Web состоит в том, чтобы зафиксировать как представление, так и содержание (то есть значение) не теряя высокоразвитую знаковую систему математики и возможность взаимосвязи.

Консорциум World Wide Web (W3C) понимал, что отсутствие основ для научной коммуникации является серьезной проблемой. В 1994 году консорциум создает рабочую группу по разработке математической разметки. Была предложена спецификация MathML, позволяющая представлять математический материал в двух видах: презентационная и содержательная разметка. Спецификация MathML поддерживает возможность преобразования между презентационной и содержательной разметками, а также в другие математические форматы, как презентационного, так и семантического вида.

Язык *MathML* является подмножеством расширенного языка разметки XML, который часто является основой для других языков. К настоящему времени консорциум W3C опубликовал 2-е издание версии 2.0 спецификации языка MathML.

Общий принцип использования MathML состоит во встраивании математических конструкций в XHTML/HTML-документ.

MathML предлагает два способа кодирования информации : *презентационный и содержательный* (см. Таблица 1)

MathML предлагает гибкую и расширяемую систему записи математического материала, позволяющей взаимодействовать с внешними программами и осуществлять высококачественное отображение в различных информационных средах

Для создания, конвертирования, просмотра, печати документов MathML предлагаются специальные редакторы формул, конвертеры и другие специализированные программные средства.

Средства отображения TeX и MathML

Существуют следующие способы просмотра математических текстов:

- Использование браузеров, которые поддерживают возможность отображения математических формул, выражений.
- Использование обычных браузеров, функциональность которых расширяется специальными дополнительными модулями, называемыми «плагинами» (plugin, add-on),

которые реализуют возможность отображения математических формул, выражений.

- Использование преобразователей из TeX и MathML в графические форматы, или в форматированный текст с графическими изображениями.

пример	презентационный	содержательный
$a - b$	<code><mrow><mi>a</mi><mo>-</mo><mi>b</mi></mrow></code>	<code><apply><minus/><ci>a</ci><ci>b</ci></apply></code>

Таблица 1. Виды разметок MathML

К первой категории относятся браузеры: Amaya, Dadzilla, Firefox или Mozilla 1.0, Netscape 7.0.

Ко второй категории можно отнести: IE 5.5+Math Player, IE 5.0 + IBM techexplorer, Netscape 6.1 + IBM techexplorer, IE 5.5 + Integre techexplorer, Netscape 4.0 + Integre techexplorer.

Браузеры

Amaya - браузер просмотра математических текстов в формате mathml, включая редактор подготовки математических текстов в этом же формате.

Версия 9.1 браузера/WYSIWYG-редактора Amaya поддерживает HTML 4.01, XHTML 1.0, XHTML Basic, XHTML 1.1, HTTP 1.1, MathML 2.0, многое из CSS 2, SVG, SMIL анимацию. Браузер позволяет просматривать и редактировать XML документы. Amaya представляет документы в структурированном виде с соблюдением DTD. Позволяется одновременно просматривать и редактировать HTML, MathML, SVG документы. Amaya позволяет расширять и модифицировать функциональность. Доступны также исходные коды этого браузера для платформ Linux, Windows (NT, XP, 2000, 98) и Mac OSX (with Fink).

Dadzilla браузер для представления арабской формы записи математических выражений с использованием MathML, которые содержат специфические символы и записываются справа налево.

Dadzilla является адаптированной версией Mozilla 1.5.1. Предполагается поддержка как латинских символов, так и арабских символов. Остается нерешенной проблема элемента `<rtl>`, который определяет RTL (right to left) направление чтения.

Плагины

Math Player является продуктом компании Design Science, быстродействующий плагин показа MathML для web-браузера Microsoft's Internet Explorer. Свободно распространяем.

Integre techexplorer Hypermedia Browser это плагин к большинству браузеров для работы с большими и расширенными математическими выражениями в Интернете. Однако возможности его несколько шире, что будет освещено ниже.

IBM techexplorer Этот плагин динамически формирует и отображает документы, которые содержат выражения TeX /LaTeX, который поддерживается только частично. Существует для платформ Windows 95, Windows NT и UNIX.

Средства создания MathML документов

- прямое программирование формул с использованием описаний на языке MathML
- использование редакторов
- использование конвертеров из TeX-формата в MathML и обратно

WebMathematica и Mathematica

Система Mathematica компании Wolfram Research была впервые выпущена 23 июня 1988 года и незамедлительно получила высочайшую оценку как научных и технических кругов Mathematica – это полностью интегрированная техническая вычислительная система. Она сочетает в себе полные вычислительные возможности и удобный интерфейс. Система включает:

- диалоговый числовой и символьный калькулятор;
- систему для визуального и звукового представления функций и данных;
- высокоуровневый язык программирования, позволяющий создавать различные программы;
- среду для моделирования, имитации и анализа данных;
- систему представления знаний в математической и технической сферах;
- язык контроля внешних программ и процессов;

- высокоуровневую оболочку для работы с файлами, текстами и данными;
- средство для создания интерактивных документов, содержащих тексты, анимационную графику и активные формулы;
- технический инструмент публикации для традиционной печати и web.

Mathematica обладает высокой интегрируемостью, широкими возможностями экспорта и импорта данных, архитектурой клиент-сервер и мощным протоколом соединения MathLink. Возможен доступ к документам клиентов извне без установки системы Mathematica. Последнее возможно выполнить через свободно распространяемый пакет MathReader, который отображает и выводит на печать мультимедийные документы, созданные в системе Mathematica.

Документы среды Mathematica можно сохранять в HTML-формате. Компания Wolfram Research была одним из участников в комитете по разработке средств поддержки математики в HTML, названных MathML. Mathematica 4 позволяет сохранять выражения в MathML. Пакет webMathematica позволяет соединить специальное ядро Mathematica и веб-страницы, напрямую подключаясь к веб-серверу. webMathematica поддерживает стандартные java технологии. Пакет webMathematica реализует технологию MSP (Mathematica Server Pages). Технология допускает включение команд из пакета *Mathematica* в HTML страницы или в формы которые помогают генерировать различные представления формул на HTML странице. Поддерживается интерактивная математика на Web-страницах. Имеется возможность конвертирования MathML выражений в изображения форматов GIF, JPEG, SVG или *Mathematica* выражения. Пакет также может выполняться конвертирование содержательного MathML в презентационный MathML, который будет корректно воспроизводим браузерами и плагинами, рекомендованными W3C.

Редакторы

EzMath

Приложение EzMath позволяет вводить математические выражения в текстовом виде на английском языке в стиле WYSIWYG - "What you see is what you get". Воспроизводит визуально и генерирует MathML содержательной разметки.

MathML Equation Editor редактор, предлагаемый фирмой Integre, позволяет выполнять описание математических выражений в виде содержательной и презентационной разметки

Abacus

Abacus - это плагин под Mozilla для написания и редактирования MathML выражений. Выражения

могут быть представлены в презентационной и содержательной разметке.

Abacus многое унаследовал от Amaya, рекомендованного W3C, применяет JavaScript и XUL для создания удобного интерфейса редактирования MathML.

MathType

MathType является профессиональной версией приложения Equation Editor, который представлен в Microsoft Office Word. MathType – это редактор для набора математических выражений, работает с текстовыми процессорами, презентационными программами, HTML-редакторами. Включает транслятор математических выражений в MathML формат.

Доступна бесплатная пробная версия на 30 дней. Есть версии под платформы Win (98SE, Me, 2000, XP), Macintosh (MacOS 9.x, OS X 10.1 and newer).

WebEQ

WebEQ – это мощный программный инструмент, который позволяет создавать Web-страницы с использованием интерактивной математики. WebEQ реализован на Java технологиях, что обеспечивает его платформу- и браузеро- независимость. Поддерживает версии для платформы Win (95, 98, 2000, Me, NT, XP), Macintosh (OS 9, OS X), Solaris, Linux. WebEQ включает модули WebEQ Editor и WebEQ Publisher, позволяющие генерировать MathML, как презентационного вида, так и содержательного. WebEQ Editor – это редактор для набора математических выражений с интуитивно понятным интерфейсом пользования. Принцип работы схож с редактором Equation Editor в Microsoft Office. Генерирует выходные данные в трех видах: презентационный и содержательный MathML, в виде графического изображения (PNG, JPG, GIF), в виде апплета. WebEQ Publisher позволяет публиковать данные в Web-формате. Входными данными могут быть MathML и WebTeX. Результатом работы могут быть опять же данные в виде MathML формата, графического изображения и апплета.

Конверторы

На сегодняшний день большая часть математических документов представлена в TeX-формате, поэтому становятся актуальными конверторы из TeX в MathML и обратно.

Конвертор из TeX/LaTeX в MathML

Разработка исследовательского центра в Онтарио по компьютерной алгебре (ORCCA). Преобразует данные из TeX в презентационный MathML и обратно. Для тестирования предлагается on-line версия. Реализован на java технологиях, что обеспечивает кроссплатформенность. Возможна настройка конвертора (mapping file). По сути mapping file – это набор шаблонов. Каждый шаблон преобразовывает

свой фрагмент TeX или MathML выражения в другой формат. Шаблон это тройка объектов, определяющая соответствия фрагментов TeX и MathML выражений между собой. Если несколько шаблонов имеют одинаковую степень соответствия, то они рассматриваются в порядке, в котором они встречаются в mapping file.

Конвертор из MathML в LaTeX

Разработка исследовательского центра в Онтарио по компьютерной алгебре (ORCCA). Пользователь может использовать несколько mapping file для определения спецификации преобразования MathML.

Конвертор из содержательного в презентационный MathML

Разработка исследовательского центра в Онтарио по компьютерной алгебре (ORCCA). При помощи XSLT-преобразования преобразует из содержательного или смешанного вида (содержательный+презентационный) MathML в презентационный вид.

Integre techexplorer

Является свободно распространяемым. Производит преобразование из TeX в презентационный MathML и обратно. Поддерживает разработку плагинов, предоставляя доступ из Java или Си++ модулей к структуре документов через Document Object Model (DOM) API.

Конвертор из MathML 2.0 в LaTeX

Проект является разработкой В. Ярошевича. Это библиотека xslt программ для конвертирования MathML документов в LaTeX документы. Производит преобразование из содержательного и презентационного MathML версии 2.0 в TeX. Продукт является свободно распространяемым. Для подключения используется стандартный синтаксис XSLT

`<xsl:import href="mmltex.xsl"/>`. Предлагается тестовая online версия.

Преобразование из содержательного MathML в TeX не всегда проходит успешно. Преобразование из презентационного MathML в TeX происходит с незначительными потерями, такими как форматирование цветом, размер символов и т.д.

TtM: конвертор из TeX в MathML, и TtH: конвертор из TeX в графический формат

TtM – это программа для конвертирования TeX или LaTeX документов в HTML+MathML документы. Для Linux платформы версия является бесплатной, для Windows платформы версия условно бесплатна.

Программа TtH генерирует HTML, конвертируя формулы TeX-документа в графические изображения, а TtM конвертирует формулы в презентационный MathML.

TtM поддерживает многое из характерных особенностей LaTeX: макросы, таблицы, библиографию. Программа не требует отдельной установки программного обеспечения TeX.

Предлагается on-line версия. Предоставление исходников программ является платным.

TeX4ht: конвертор из TeX /LaTeX в MathML

TeX4ht (реализован на языке C) конвертирует TeX-документ в HTML-документ с картинками. Другая компонента TeX4moz конвертирует из TeX-документа в XHTML+MathML документ. Предложенный для просмотра тестовый xml-документ содержит MathML презентационной разметки. Обе компоненты устанавливаются под пакет программы MicTex.

Средства генерирования графических форматов по математическим выражениям.

Конверторы из TeX формата в графический формат

Textogif – это программа, реализованная на Perl для преобразования выражений в небольших LaTeX-файлах в картинки. При работе обращается к средствам TeX.

Mimetex – CGI-программа, реализованная на C. На вход принимает TeX-формулу. В web-странице формула указывается в качестве параметра URL-обращения к программе (значение атрибута src тега img).

Например, ``

MathML и SVG

SVG – это язык для описания двумерной графики и графических приложений на основе стандарта XML SVG 1.1, рекомендован консорциумом W3C. Для просмотра SVG –формата необходимо установить соответствующий плагин к браузеру. Существует много коммерческих программных продуктов конвертирования из MathML в SVG формат: Mathematica, WeBEQ Publisher.

Заключение

При работе с математическими данными мы сталкиваемся с двумя проблемами: просмотр и создание документа с математическими выражениями. Некоторые браузеры позволяют просматривать математические выражения без дополнительных компонент. Для другой части браузеров необходима установка «плагинов».

Для представления математического текста в сети рекомендуется использовать язык математической разметки MathML. Получить MathML позволяют многочисленные редакторы или браузеры-редакторы.

В силу исторически сложившихся условий огромное количество математических документов представлено в TeX формате, вследствие чего приобретают актуальность конверторы из TeX в MathML и обратно.

Стоит учитывать тот факт, что большая часть пользователей Интернета будут просматривать документы с математическими выражениями, не прибегая к помощи плагинов. Поэтому остается актуальным генерирование HTML документов с +GIF (JPEG, PNG) изображениями, что может быть сделано конвертированием из TeX/LaTeX или MathML форматов в графический формат. О свободной поддержке конвертирования из MathML в SVG было заявлено фирмой ADOBE, но, к сожалению, это пока не реализовано.

Как упоминалось, существует два вида представления информации MathML: презентационный и содержательный. Опробованные программные продукты позволяют сгенерировать презентационный MathML.

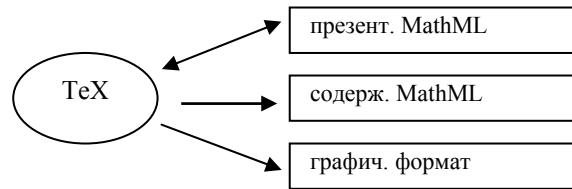
Как показывают тестовые on-line версии, наиболее точные решения преобразования из MathML в TeX и обратно предлагают коммерческие программные продукты, свободно распространяемые программные продукты не дают стабильного решения.

На основе свободно распространяемых продуктов можно получить следующий цикл преобразований:

Пакет Mathematica позволяет в ряде случаев генерировать содержательный MathML.

Для публикации математических данных в сети возможно создание документов вида

HTML+графическое изображение и HTML+MathML.



На стороне веб-сервера математические выражения необходимо поддерживать как в графическом формате, так и в формате MathML. Из рассмотренного программного обеспечения в статье можно выстроить следующую цепочку:

Конвертор из TeX/LaTeX в MathML (ORCCA), Конвертор из MathML в LaTeX (ORCCA), Конвертор из содержательного в презентационный MathML (ORCCA), Mimetex(CGI-программа).

Линейка программ, являющихся разработкой центра ORCCA, выполнена на технологиях JAVA и XSLT, которые позволяют поддерживать принцип интероперабельности информационных систем.

На сегодняшний день стабильного решения из свободно распространяемых продуктов по генерированию графического формата из MathML не существует, возможно генерирование графического формата из TeX-формата, что реализовано в Mimetex.

Сводная таблица инструментов представления и конвертирования математических данных

Программные продукты	Редактор (презентационный) MathML	Редактор (содержательный) MathML	Из TeX в презентационный MathML	Из презентационного MathML в TeX	Из содержательного MathML в TeX	Из содержательного MathML презентационный MathML	Из TeX в графический формат (GIF, JPG, PNG)	Из MathML в MathML
Mathematica WebMathematica	+	+	+	+	+	+	+	+
Amaya	+	+						
ExMath	+	+						
MathML Equation Editor	+	+						
MathType	+	+						
WebEQ	+	+					+	
Конвертор из TeX/LaTeX в MathML (ORCCA)			+					
Конвертор из MathML в LaTeX (ORCCA)				+				
Конвертор из содержательного в презентационный MathML (ORCCA)						+		

Integre techexplorer Hypermedia Browser			+	+			+	
Конвертор из MathML 2.0 в LaTeX Дружеская				+	+			
TtM конвертор из TEX в MathML				+				
TTH конвертор из TEX в графический формат							+	
TeX4ht конвертор из TeX/LaTeX в MathML				+	+			
TeXlogit							+	
Mimeticx							+	
<p>⋮ - коммерческие программные продукты</p> <p>▨ - условно бесплатные программные продукты</p> <p>▩ - свободно распространяемые программные продукты</p>								

Список используемых источников

- Add Math to Web pages with HP EzMath v1.1
<http://www.w3.org/People/Raggett/EzMath/>
- Gurari E. TeX4ht: LaTeX and TeX for Hypertext
<http://www.cse.ohio-state.edu/~gurari/TeX4ht/mn.html>
- http://www.elbook.bsu.by/OurMath/Library/Articles/from_wolfram.htm (WebMathematica)
- IBM techexplorer Hypermedia Browser
<http://tex.loria.fr/outils/techexplorer.pdf>
- Integre techexplorerTM Hypermedia Browser
<http://www.integretechpub.com/techexplorer/>
- Kennedy J. Technical Report for the Mathematically Inclined on the MathML Internet Markup Language and the Unicode Character Standard
http://homepage.smc.edu/kennedy_john/JKMATHMLTECHREPORT.PDF#search='Pavi%20Sandhu%20XML%20and%20Mathematica'
- Mathematical Markup Language (MathML) Version 2.0 <http://www.w3.org/TR/2003/REC-MathML2-20031021/>
- MathPlayer
<http://www.dessci.com/en/products/mathplayer/>
- MathType
<http://www.dessci.com/en/products/mathtype/>
- Projects involving MathML in progress
<http://www.w3.org/Math/iandi/interop02.html#intro>
- Smirnova E., Watt S. MathML to TeX Conversion: Conserving high-level semantics
<http://www.mathmlconference.org/2002/presentation/smirnova/#1>
- SVG Conversion Tools
<http://www.w3.org/Graphics/SVG/SVG-Implementations.htm#convert>
- TeX/LaTeX to MathML Online Translator
<http://www.orcca.on.ca/MathML/texmml/textomml.html>
- The W3C MathML software list
<http://www.w3.org/Math/Software/>
- TtM, a TeX to MathML translator
<http://hutchinson.belmont.ma.us/tth/mml/>
- WebEQ
<http://www.dessci.com/en/products/webeq/>
- Welcome to Amaya W3C's Editor/Browser
<http://www.w3.org/Amaya>
- Wickham-Jones T. webMathematica: How to Deliver Computational and Visualization Services from a Web Server
<http://www.mathmlconference.org/2002/presentation/twj/>
- XSLT MathML Library
<http://sourceforge.net/projects/xsltml/>
- Гурский Ю. Работа с форматом SVG в Adobe Illustrator 10 <http://www.hardline.ru/3/37/3381/>
- Система Mathematica и базы данных
<http://www.wolfram.com/products/mathematica/index.html>

Review the tool of the mathematics data

In article are presented famous tools for work with mathematics data and various way presentation in

Web: TeX/LaTeX, MathML, various graphical formats too. The main functions, capacity, merits and demerits of tools are discussed; also is presented comparative analysis of conversion mathematics data, presented in TeX/LaTeX and MathML formats.