



OSTIS-2011

(Open Semantic Technologies for Intelligent Systems)

УДК 004.9:510

WEB-ОРИЕНТИРОВАННАЯ НОТАЦИЯ МАТЕМАТИЧЕСКИХ ТЕКСТОВ, СОХРАНЯЮЩАЯ СЕМАНТИКУ

А.И. Вовк (vovk@ndiasb.kiev.ua)

*Государственный научно-исследовательский институт
автоматизированных систем в строительстве, г.Киев, Украина*

Д.А. Гирнык (den@girnyk.com)

*Международный научно-учебный Центр информационных технологий и систем
Национальной Академии наук Украины, г.Киев, Украина*

В работе рассмотрена WEB-ориентированная нотация, реализующая натуральный язык представления математических текстов и сохраняющая их семантику. Разработаны средства визуализации нотации для браузеров и экспорта в стандартные форматы Latex, MathML и PDF. Предложенная WEB-нотация обеспечивает совместимость, интероперабельность и интегрируемость подготовленных текстов с другими приложениями.

Ключевые слова: математическая нотация, математические формулы, WEB-математика.

Введение

Особенностью существенной части многих научных и образовательных ресурсов, содержащих математику, физику, экономические науки и т.п., является наличие специфических естественнонаучных текстов – математических формул. Математические формулы в виртуальном образовательном процессе могут быть использованы более широко, чем просто в электронных текстах. Это могут быть форумы, чаты, другие формы обмена данными в электронном виде, электронные библиотеки, специализированные практикумы по естественным наукам, тесты и т.д., а также приложения, связанные с электронной обработкой математических формул [Вовк и др., 2007].

Под термином нотация (от лат. Notatio – записи, обозначения) понимают систему условных обозначений, принятых в конкретной области знаний или деятельности. Нотация содержит множество символов, используемых для представления понятий и их взаимоотношений, которые образуют алфавит нотации, а также правила их применения.

1. Компьютеризованная нотация

Исторически на протяжении достаточно большого периода 18-го и 19-го веков сформировалась стандартная нотация математических формул, используемая сегодня на бумажных носителях в виде разметки на плоскости букв разных алфавитов, специальных математических символов и символических выражений, которые должны иметь точное семантическое значение. Решающий вклад в создание математической нотации внесли Эйлер, Джонс, Кантор, Валлис, Гаусс, Лейбниц и другие математики.

С наступлением информационной эры появилась компьютеризованная нотация математических объектов. Сначала это был синтаксис арифметических выражений первых языков программирования (Fortran, Algol, Basic, Pascal), специализированных математических пакетов (Maxima, Maple, MathCAD), тестовых редакторов, а в дальнейшем были разработаны стандарты языков разметки текста (HTML, MathML, TeX и др.).

Характерной особенностью применений в программировании является строчное представление формул, определяется последовательностью их обработки. Нотация языков программирования сохраняет семантику математических формул, приспособленная для вычислений по формулам, является инструментом всех программистов, но не воспринимается людьми, далекими от программирования, не принимается журналами к печати и т.д.

Со временем появились визуальные редакторы для ввода математических формул. На сегодня они используются практически во всех Desktop-пакетах по математике.

Под обработкой математических текстов в широком смысле понимают всю цепочку: введение математических формул пользователем, хранение формул, собственно обработка формул, отображение формул в рамках единой технологии.

2. Гипертекстовая разметка

Для разметки текстов журнальных статей широко используется нотация TeX, разработанная еще Д. Кнудом [Knuth, 1984]. Сегодня издается, по разным оценкам, от 20 до 30 тысяч электронных научных журналов, только IEEE ежегодно публикует более полумиллиона страниц технических работ. Все крупнейшие издательства публикуют свои журналы в электронном виде. Появилось значительное количество электронных журналов открытого доступа. Например, электронный архив Science Direct включает около 2,5 тыс. полнотекстовых электронных журналов, содержащих более 8 млн. статей.

К сожалению, при использовании TeX теряется информация о семантике математических выражений. Однако такой формат предоставления текстов настолько распространился, что ведущие научные журналы мира уже не принимают ни для редактирования, ни даже для ознакомления рукописи в других форматах, например, в формате Microsoft Word, поскольку отражение его формул несовместимо с TeX.

Одним из первых поднял вопрос о средствах работы с математическими текстами в Интернет К.Фостер [Foster, 1999].

На сегодня классическим средой работы с естественнонаучными текстами в Интернете является язык MathML. Нотация MathML реализует одну из основных тенденций современных информационных технологий - отделение контента от формы представления. MathML структурирует документ в хорошо понятных автору терминах содержательного (логического) уровня - теорема, определение, формула, раздел т.д. Это предоставляет широкие возможности для многоуровневого структурирования данных и расширенного поиска, обеспечивая одновременно гибкость стилей оформления.

MathML является открытым стандартом, поддерживаемым консорциумом W3, что очень важно для построения сложных систем, ориентированных на использование широкими массами пользователей. В своей «презентационной» части MathML идеологически очень близок к TeX-формату.

Существуют конвертеры для преобразования текстов из MathML в TeX (LaTeX) для печатных изданий.

В техническом комитете Unicode и соответствующем органе, контролирующем универсальный набор символов ISO 10646, уже рассмотрен проект о размещении в кодовых таблицах новых символов для математики [Unicode, 2010].

В настоящее время ведутся активные исследования в области методов, систем и средств использования математических WEB-представлений в интернете для интерактивного обучения математике [Yang, 2010], для тестирования по естественнонаучным дисциплинам [Yue-sheng, 2008], для обучения и подготовке по математическим и статистическим дисциплинам на базе "облачных" технологий [Sousse, 2010] и др. смежных областях. Попытка обзора последних решений в области обучения математики предпринята в работе [Chaamwe, 2010].

В последнее время все большую популярность приобретают средства динамического интерактивного общения в Интернете: форумы, чаты, системы мгновенного обмена данными (ICQ, Skype) и т.д. Эти средства с успехом могут быть использованы в общении научных работников и в дистанционном обучении по естественнонаучным предметам, если содержат поддержку математических формул.

3. Натуральная нотация

По умолчанию в современных браузерах реализовано строчное введение данных, поэтому и математические формулы вынуждены вводиться в одну строку. Но, поскольку даже в общеобразовательных школах изучается как минимум один язык программирования (Pascal), поддерживающий строчную нотацию формул, такое ограничение не кажется критичным.

Для интерактивного общения в Интернете прежде всего актуальны следующие требования к математической нотации:

- возможность однострочного набора математических формул;
- лаконичность нотации;
- скорость набора формул;
- "естественность", "привычность" строчного набора;
- сохранение семантики математических выражений;
- кроссплатформенность;
- интероперабельность;
- приспособленность к интеграции с другими приложениями.

Для контроля формул и комфортной работы с ними желательно, а во многих случаях и необходимо, иметь их изображение плоскости, например, в отдельном окне визуализации. Важно обеспечить динамичность корректировки ввода формул. Реакцию системы для этой операции желательно выдержать минимальной.

Однако большую интерактивность обеспечивают встроенные средства, плагины к браузерам (математические плееры, ActiveX-приложения и т.д.). Математические плееры MathML по умолчанию встроены в браузер FireFox и доступны для Internet Explorer.

Еще в 2000 году была предложена удовлетворяющая указанным требованиям натуральная нотация для гипертекстовой разметки [Вовк, 2000], структура которой максимально приближена к способу отображения математических выражений в языках программирования.

Предложенная нотация позволила в определенной степени решить одну из наиболее серьезных проблем - введение и корректировку формул. Еще в процессе разработки MathML стало ясно, что требования к синтаксису ввода настолько разные, что не может быть одного метода, удовлетворяющего всех пользователей. Многие из участников рабочей группы разработали форматы ввода отдельно для своих инструментов. Рабочая группа W3C предоставляет технические рекомендации для тех, кто занимается подобными разработками. Поэтому синтаксис ввода не является частью основной рекомендации на MathML [W3C, 2010].

Создан редактор MathTextView (MTV) для реализации натуральной нотации. Он сохраняет как синтаксис формул, так и ее семантику. Имеется принципиальная возможность превращения нотации в бескомпромиссную текстовую запись, идеально подходящую к машинной обработке математических выражений. Допускается возможность семантического контроля нотации.

Высокая скорость освоения редактора обеспечивается знанием пользователями языков программирования Pascal, Basic, C, в которых математические выражения записываются аналогично.

На сегодня редактор MTV реализован как плагин ActiveX для браузера Internet Explorer.

Кроссбраузерно преобразование нотации введенных формул в классическую плоскостную нотацию может быть осуществлено путем обработки на Интернет-сервере с последующим получением изображения в виде рисунка (png, gif, jpg). Сайты таких преобразований доступны, как на коммерческой основе, так и свободно, например, [MathTran, 2006].

Разработана технология серверной обработки информации в формате MathTextView. Конечное графическое представление выполнено с помощью стандартного формата DVI (DeVice Independent), что дает возможность дальнейшего преобразования во все форматы (PNG, PDF, TeX и др.) стандартными программными средствами.

Таким образом, сохраняется семантика математических формул и есть возможность публикации в форматах всех издательских систем. Поддержка семантики позволяет проводить электронную обработку выражений прямо в Интернете. Например, на сайте представлен калькулятор интерактивного дифференцирования [MathTextView, 2000].

На рисунке приведен мгновенный снимок on-line WEB-дифференцирования на сайте MathTextView.

В верхней строке отображается текст в натуральной нотации MTV, в поле ниже – интерактивная визуализация текущего математического выражения в браузере. Курсором и прямоугольником отображено место, где продолжается ввод символов, в нотации и в интерактивной визуализации соответственно.

$$\begin{aligned}
 & \text{\$ \$ } df(x^x/a(x)+1/(x+\sin(x+\square)).x)=d? \text{\$ \$} \\
 & \frac{d\left(\frac{x^x}{a(x)} + \frac{1}{x+\sin(x+\square)}\right)}{dx} = \frac{(x^x \cdot \ln(x) + x^x) \cdot a'(x) - x^x \cdot a''(x)}{a^2(x)} - \frac{1}{(x+\sin(x+\square))^2} \cdot (1 + \cos(x+\square))
 \end{aligned}$$

Рисунок 1 – Скриншот браузера страницы символьного дифференцирования на сайте MTV

Нотация MathTextView совместима с языками программирования и позволяет напрямую обмениваться с компьютерными программами обработки формул в исходных кодах.

Такую же нотацию используют многие пакеты по символьной WEB-математике. Для примера, покажем обработку формулы MTV на ресурсе символьного интегрирования Wolfram.

Библиографический список

[Вовк и др., 2007] Вовк А.И., Гирнык Д.А. Язык общения математиков в Интернете. В кн.: New Information Technologies in Education for all: State of the art and Prospects (ITEA-2007), Kiev, Ukraine, IRTС, 21-23 November 2007, p.p. 96 – 103.

[Knuth, 1984] Donald E. Knuth, The TeXbook (Reading, Massachusetts: Addison-Wesley), 1984.

[Foster, 1999] Foster K.R. "Math on the Internet". IEEE Spectrum, Vol 36(4):, Apr 1999. p.p. 36-40.

[Unicode, 2010] Unicode Entity Codes for Math [Электронный ресурс]. – 2010. - Режим доступа: <http://tlt.its.psu.edu/suggestions/international/bylanguage/mathchart.html>. – Дата доступа: 03.09.210.

[Yang, 2010] Li Yang, Mo Qian, Wang Fang. An Interactive Mathematics Education Platform Based on Topic-Based Deep Search. Second International Workshop on Education Technology and Computer Science (ETCS), 6-7 March 2010, Vol: 2, p.p. 163 – 169.

[Yue-sheng, 2008] Gu Yue-sheng, Zhu Jia-yi. Uploading Strategy of the Formula in the Web-Based Mathematics Testing System. International Conference on Computer Science and Software Engineering, 12-14 Dec. 2008, Vol: 5, p.p. 624 – 626.

[Sousse, 2010] Sousse T. Learning Math and Statistics on the Cloud, Towards an EC2-Based Google Docs-like Portal for Teaching / Learning Collaboratively with R and Scilab. 2010 10th IEEE International Conference on Advanced Learning Technologies. July 05- 07 2010, pp. 752-753.

[Chaamwe, 2010] Chaamwe N. Integrating ICTs in the Teaching and Learning of Mathematics: An Overview. Second International Workshop on Education Technology and Computer Science (ETCS), 6-7 March 2010, Vol: 2, p.p. 397 – 400.

[Вовк, 2000] Вовк А.И. и др. Язык представления математических текстов в интернете. Теорія та методика навчання фундаментальних дисциплін у вищій школі. Збірник наукових праць. Кривий Ріг, Вид. відділ НМетАУ, 2004.

[W3C, 2010] World Wide Web Consortium (W3C) [Электронный ресурс]. – 2010. - Режим доступа: <http://www.w3c.org>. – Дата доступа: 30.09.2010.

[MathTran, 2006] MathTran (транслятор математических нотаций) [Электронный ресурс]. – 2010. - Режим доступа: <http://www.mathtran.org>. – Дата доступа: 14.09.2009.

[MathTextView, 2000] Математика в Интернете [Электронный ресурс]. – 2000. - Режим доступа: <http://math.accent.kiev.ua>. – Дата доступа: 30.11.2010.