

# ОРГАНИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ ЯЗЫКОВ РАЗМЕТКИ В СИСТЕМЕ АВТОМАТИЗАЦИИ ЭЛЕКТРОННЫХ ХРАНИЛИЩ: СЕМАНТИЧЕСКИЙ ПОДХОД

А.М. Елизаров, Е.К. Липачев, М.А. Малахальцев

Технологии семантического веба предлагают стандартную процедуру создания языка разметки, адаптированного к определенной предметной области [1], [2].

К настоящему времени созданы специализированные языки разметки для большинства естественнонаучных предметных областей, широко применяемые в научной среде: Mathematical Markup Language (MathML), Chemical Markup Language (CML), Geography Markup Language (GML), Finite Element Modeling Markup Language (femML), Mathematics Education Markup Language (MeML), Materials Markup Language (MatML), Numerical Data Markup Language (NDML), Relational-Functional Markup Language (RFML), Robotic Markup Language (RoboML).

Например, с помощью CML записывается информация о молекулярных структурах, химических реакциях, спектрах, неорганических кристаллах и объектах квантовой химии. Важно отметить, что файлы, содержащие разметку согласно правилам CML, являются XML-файлами, и поэтому для их создания и обработки можно использовать специальные программные средства [2], [3]. Этот язык активно развивается в течение последних десяти лет, появляются его диалекты, например, Polymer Markup Language (PML) [4]. Отметим, что язык разметки CML разработан как часть проекта Open Molecule Foundation.

Для представления математических формул в рамках семантического веба используется язык разметки MathML, разработка которого ведется консорциумом W3C (<http://www.w3c.org>) с 1998 года. MathML является стандартом представления математической информации в электронной форме [5], [6]. Для хранения и электронного обмена математическими моделями применяется язык CellML (Cell Markup Language). Он широко применяется в биологическом моделировании. Отметим, что CellML поддерживает спецификацию MathML. Для описания свойств материалов можно использовать язык Materials Markup Language (MatML). Географическим сообществом используется язык Geography Markup Language (GML).

Наряду со специализированными языками разметки, предназначенными для описания результатов, активно создаются языки разметки для оформления публикаций, например, языки разметки диссертаций: Dissertation Markup Language (DiML) [8], Electronic Thesis and Dissertation Markup Language (ETD-ML) [9]. Разработаны также языки для электронных публикаций, например, A markup language for scientific, technical and medical publishing (STMML) [10], для электронных учебников, например, Tutorial Markup Language (TML) [11].

Каждый язык разметки, согласно стандартам и рекомендациям консорциума W3C, формируется закреплением множества элементов и их атрибутов путем создания XML-схемы или DTD-описания, разработкой программных средств преобразования и отображения документов, написанных на этом языке, и созданием спецификации данного языка разметки.

В настоящее время во всем мире, в том числе и в России, идет активная работа по созданию электронных хранилищ, в частности, создаются и развиваются разнообразные коллекции электронных научных документов. Разработаны основные принципы организации таких коллекций и соответствующее программное обеспечение. По этим принципам работает большинство электронных хранилищ, например, Научная электронная библиотека eLibrary.ru (см. [7]). В большинстве случаев научная электронная коллекция представляет собой набор документов (в основном текстов статей и книг) и их <библиографических> описаний на основе XML. Электронные коллекции уже сейчас позволяют организовать поиск не только по текстовой информации, но и по XML-описанию.

В электронных хранилищах, содержащих научную информацию, относящуюся к различным предметным областям, должна быть создана система согласования и управления междисциплинарным контентом. Имеющиеся на сегодняшний день технологии семантического веба, в частности, разработанные языки разметки естественнонаучного контента, позволяют решить эту задачу.

Ряд подходов к решению задач управления специализированными научными ресурсами был разработан нами в рамках нескольких проектов РФФИ, в частности, эти подходы были реализованы в программной среде электронного научного журнала. Они основаны на использовании XML-схем, DTD-описаний и XSLT-преобразований для управления информационными потоками. Примером такого решения является автоматизация работы электронного журнала <Lobachevskii Journal of Mathematics> (<http://ljm.ksu.ru>) [7].

В настоящее время ведется разработка программной среды, позволяющей осуществлять управление междисциплинарным контентом. Основная цель - сведение в единую схему основных специализированных языков разметки научно-предметных областей и обеспечение их взаимодействия в рамках единой программной среды. Это обеспечит возможность решения ряда принципиальных задач, возникающих при построении масштабных электронных хранилищ естественнонаучного контента, позволит обрабатывать в едином программном комплексе разнообразные типы документов (научные статьи, результаты наблюдений и

экспериментов, программные продукты), которые представлены в различных форматах, используемых в научной среде. Одним из основных элементов этой системы является интеллектуальный поиск, позволяющий пользователю отслеживать динамические связи между документами с учетом специфики предметной области. В соответствии с принципами семантического веба разрабатываемая система имеет интерфейс, приспособленный для взаимодействия на программном уровне, т. е. соответствует концепции <машинно-ориентированного подхода> [1].

Работа выполнена при финансовой поддержке РФФИ (проект 09-07-12059офи\_м) и РГНФ (проект 07-01-12146).

#### ЛИТЕРАТУРА:

1. Berners-Lee T. Semantic Web Road Map. - <http://www.w3.org/DesignIssues/Semantic.html>; Рус. перевод: <http://gridclub.ru/library/publication.2007-04-23.2195467714/view>
2. Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Языки разметки семантического веба. Практические аспекты. - [http://www.ksu.ru/fpk/docs/lip\\_mal.pdf](http://www.ksu.ru/fpk/docs/lip_mal.pdf)
3. Murray-Rust P., Rzepa H.S. Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles// J. Chem. Inf. Comput. Sci. - 1999. - V. 39. - P. 928-942.
4. Adams N., Winter J., Murray-Rust P., Rzepa H.S. Chemical Markup, XML, and the Worldwide Web. 8. Polymer Markup Language//J. Chem. Inf. Model. - 2008. - V. 48. - P. 2118-2128.
5. Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Основы MathML. Представление математических текстов в Internet. Практическое руководство. - Казань: Изд-во Казанского математического общества, 2004. - 60 с.
6. Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Основы MathML. Представление математических текстов в Internet. - Казань, 2008. - 101 с. - <http://www.niimm.ksu.ru/data/preprints/>
7. Веселаго В.Г., Елизаров А.М., Липачёв Е.К., Малахальцев М.А. Формирование и поддержка физико-математических электронных научных изданий: переход на технологии семантического веба// В кн. <Научно-исследовательский институт математики и механики им. Н.Г. Чеботарева Казанского государственного университета. 2003 - 2007 гг.>. Кол. монография под ред. А.М. Елизарова. - Казань: Изд-во Казан. ун-та, 2008. - С. 456-476.
8. Documentent und Publicationsserver, Humbolt University. - <http://edoc.hu-berlin.de/diml/>
9. Kipp N. A. Electronic Thesis and Dissertation Markup Language (ETD-ML) User's Guide. - <http://etd.vt.edu/etd-ml/userguid.htm>
10. Murray-Rust P., Rzepa H.S. STMML. A Markup Language for Scientific, Technical and Medical Publishing. - <http://www.ch.ic.ac.uk/rzepa/codata2/>
11. TML Language Specification. - <http://www.ilrt.bris.ac.uk/netquest/about/lang/>