

ТЕХНОЛОГИИ XML И XML-ДАННЫЕ

М.Р. Когаловский

Институт проблем рынка РАН, Москва

e-mail: kogalov@cemi.rssi.ru

*Базы данных и информационные технологии XXI века.
Материалы международной научной конференции,
Москва, 29-30 сентября 2003 г. – М.: РГГУ, 2004.*

Абстракт

Вторая половина 90-х годов стала временем рождения новых технологий управления данными в среде Веб, основанных на разработанном консорциумом W3C стандарте расширяемого языка разметки XML. Их создание направлено, прежде всего, на преодоление ограничений существующей версии Веб и разработку технологической платформы Веб нового поколения, называемого семантическим Веб, на обеспечение возможности его развития на длительную перспективу.

Функциональное ядро новой технологической платформы Веб, называемой в докладе *платформой XML*, составляет комплекс стандартов консорциума W3C, основанных на языке XML. Ряд из них быстро завоевал широкое признание. Сфера их применения стала довольно широкой и выходит за рамки непосредственных потребностей Веб. Они оказывают заметное влияние на развитие стандартов в других областях информационных технологий как в горизонтальной, так и в вертикальной сферах, в частности, в технологиях баз данных, хранилищ данных, CASE-технологиях, в области интеграции данных, в области электронных библиотек, электронного бизнеса, им отводится важное место в технологиях электронного правительства. Активно развиваются тенденции интеграции технологий XML с другими пластами информационных технологий. Технологии XML с полным правом можно считать информационными технологиями 21 века.

В докладе анализируются предпосылки создания платформы XML и ее истоки, дается краткий обзор функциональных возможностей комплекса стандартов, составляющего эту платформу, и стандартов ее окружения, обсуждается концепция расширяемости, обозначенная в названии языка XML.

Центральное место в докладе занимает рассмотрение особенностей XML-данных и возможностей их моделирования, обеспечиваемых стандартами W3C, сопоставление подходов к управлению данными XML с подходами, принятыми в технологиях традиционных баз данных. Обсуждаются, наконец, вопросы конвергенции и интеграции технологий XML и технологий баз данных.

Эта работа частично поддержана грантами РФФИ 01-07-90444 и РГНФ 03-02-12008.

1. Введение

Прошло лишь около десяти лет со времени создания системы, которую называют Всемирной паутиной (World Wide Web, WWW или Веб). За сравнительно короткий период своего существования эта глобальная распределенная гипермедийная информационная система впитала в себя гигантские объемы информационных ресурсов, обрела многие сотни миллионов пользователей на всех континентах. Эта глобальная информационная система интенсивно вторгается в другие области информационных технологий, стала одним из важных звеньев инфраструктуры информационного общества.

Такая популярность Веб обеспечивается многими его достоинствами: открытым характером системы, демократичной ее организацией, привычной архитектурой «клиент–сервер», возможностью свободного доступа к большинству ресурсов системы без ограничений по времени, наличием различных средств свободно распространяемого программного обеспечения как клиентского, так и серверного, прозрачностью распределенности информационных ресурсов Веб для пользователя, независимостью используемых технологий

от аппаратно-программных платформ, простотой языка разметки HTML, легкостью подготовки и публикации документов, простой техникой идентификации информационных ресурсов, низким уровнем требований к квалификации пользователя и к конфигурации его технических средств, простым клиентским интерфейсом с возможностью доступа к информационным ресурсам с помощью естественной навигации по их структуре и т.д.

Немаловажным фактором, определяющим чрезвычайно высокую социально-экономическую значимость Веб, служит также то обстоятельство, что средой его обитания является глобальная коммуникационная среда Интернет, обеспечивающая теледоступ пользователей и благоприятные возможности для интеграции распределенных информационных и вычислительных ресурсов. Благодаря этому Веб – не только гигантская информационная система, но и эффективная платформа для разработки многочисленных новых приложений и технологий – электронных библиотек, систем электронного бизнеса, виртуальных предприятий, крупных корпоративных информационных систем.

Однако количество переходит в качество. Уже в первые годы триумфального роста Веб проявился ряд ограничений, свойственных используемым в нем технологиям, которые стали сдерживать дальнейшее его развитие. Достоинства стали оборачиваться недостатками.

Вторая половина 90-х годов прошедшего века стала временем радикальных перемен в технологиях Веб. Новые подходы в области технологий Веб, которые начали конструктивно воплощаться в жизнь на пороге начинающегося века, направлены, прежде всего, на преодоление этих ограничений и создание новой технологической платформы, способной обеспечить потенциал для создания Веб нового поколения, называемого семантическим Веб [19, 20], и возможности его развития на длительную перспективу. основополагающую роль в технологическом переоснащении Веб стал играть новый язык разметки XML, который разработан консорциумом W3C [32], ответственным за техническую политику развития Веб.

Новая технологическая платформа Веб, образованная стандартом XML и рядом других основанных на этом языке стандартов, быстро завоевала широкое признание. Сфера применения стандартов XML интенсивно расширяется и вышла за рамки непосредственных потребностей Веб. Они оказывают заметное влияние на развитие стандартов других областей информационных технологий, в частности, технологий баз данных, хранилищ данных, интеграции данных, CASE-технологий. В свою очередь, технологии баз данных оказывают значительное влияние на развитие стандартов XML. Вместе с тем, стандарты XML стали основой разработки новых подходов в области электронного бизнеса, им отводится важное место в технологиях электронного правительства. Активно развиваются тенденции интеграции технологий XML с другими пластами информационных технологий. Технологии XML с полным правом можно считать информационными технологиями 21 века.

В предлагаемом докладе анализируются предпосылки рождения технологий XML и их истоки, дается краткий обзор функциональных возможностей комплекса стандартов XML, исследуются особенности XML-данных и их моделирования в стандартах W3C, сопоставляются подходы к управлению данными XML с подходами, принятыми в технологиях традиционных баз данных. Рассматриваются также вопросы конвергенции и интеграции технологий XML и технологий баз данных.

Более подробное обсуждение принципов организации платформы XML и составляющих ее стандартов, их назначения, принципов моделирования данных, принятых в стандартах XML, направлений использования стандартов этой платформы в горизонтальной и вертикальной сферах, основных подходов к разработке XML-ориентированных СУБД и других вопросов, связанных с этой новой технологической платформой Веб, можно найти в наших работах [3-12].

2. Предпосылки рождения технологий XML

Создание Веб безусловно можно считать одним из крупнейших научно-технических достижений последнего десятилетия XX века. Благодаря реализации этого проекта рождается целый ряд новых информационных технологий, имеющих весьма значимые социально-экономические последствия. В короткие сроки Веб стал беспрецедентно интенсивно развивающейся глобальной открытой бесконечно масштабируемой распределенной

гипермедийной системой. Количество пользователей и объем представленных в ней информационных ресурсов продолжают чрезвычайно быстро наращиваться.

Вместе с тем, за несколько лет интенсивного развития потенциал качественного совершенствования технологий Веб-1 оказался в значительной мере исчерпанным.

Сдерживающее влияние на дальнейшую эволюцию технологий Веб и на расширение сфер их применения стали оказывать, прежде всего, слабые стороны языка HTML - основного выразительного и структурообразующего средства представленных в Веб гипермедийных информационных ресурсов, а также ограниченные функциональные возможности среды поддержки этого языка в Веб. В языке HTML имеются лишь весьма скромные средства представления метаданных, описывающих содержание информационных ресурсов Веб-серверов (тег META). Следствием этого является отсутствие возможности верификации логической целостности данных, содержащихся в HTML-страницах, высокий уровень информационного шума в результатах обработки пользовательских запросов поисковыми машинами Веб. Ряд проблем связан с идентификацией страниц Веб и составляющих их ресурсов по местоположению содержащих их файлов в Интернет. Протокол доступа HTTP не обеспечивает удаленного вызова процедур, что значительно ограничивает возможности разработки Веб-приложений.

Огромный объем накопленных в среде Веб информационных ресурсов требует для эффективного их использования новых подходов, позволяющих возложить на вычислительную технику значительно большую, чем в действующем Веб, нагрузку по их анализу, обработке, поиску и интерпретации содержания.

Разработка и конструктивное воплощение в виде комплекса взаимосвязанных стандартов радикально новых подходов, которые обеспечат дальнейшее развитие Веб, стали важнейшей стратегической задачей консорциума W3C, ответственного за техническую политику развития технологий Веб. Создание расширяемого языка разметки XML и основанного на этом языке комплекса других стандартов, составляющих технологическую платформу Веб нового поколения, являются важным результатом деятельности консорциума. Работы по формированию этой модульной платформы, которую мы называем *платформой XML*, пока не завершены, хотя они ведутся довольно интенсивно.

3. Истоки технологий XML

Технологии XML создавались отнюдь не «с чистого листа». В них обеспечена преемственность с Веб первого поколения. Используются некоторые его ключевые подходы, например, организация информационных ресурсов в виде гипертекста, навигационный доступ к информационным ресурсам, взаимодействие пользователей со средствами управления информационными ресурсами Веб на основе архитектурного принципа «клиент-сервер». Наряду с новыми возможностями сохранена также идентификация информационных ресурсов по месту их хранения в Интернет. С архитектурной точки зрения, новые средства Веб по-прежнему относятся к прикладному уровню в семиуровневой эталонной модели взаимосвязи открытых систем (OSI).

Однако для обеспечения новых возможностей потребовались и другие решения. Ряд из них заимствован или, по крайней мере, уже использовался ранее в технологиях баз данных. В спецификациях стандартов XML существенное место занимают такие понятия, рожденные в области баз данных, как модель данных, схема, язык запросов, самоописываемость данных. Концепция схемы базы данных, отчужденной от самих данных, была введена еще CODASYL (1969). Принцип *самоописываемости баз данных* был впервые использован в реляционной СУБД проекта MacAims (1970), а впоследствии узаконен в стандартах языка SQL. Интересно заметить, что основная структурная единица XML-данных, называемая *XML-документом*, имеет иерархическую структуру и включает в качестве элементов повторяющиеся группы с переменным числом экземпляров. В точности такую структуру имеет запись базы данных в модели данных CODASYL.

Создатели XML остались на позициях подхода к представлению информационных ресурсов Веб в форме *гипертекста*, идеи которого возникли еще в конце 40-х годов и начали практически использоваться в 60-е годы. При этом для определения гипертекстовой структуры стали использовать *языки разметки*, одним из ранних представителей которых является

разработанный американским математиком Д. Кнудом в 70-е гг. язык TeX. Авторы новой технологической платформы вновь обратились к истокам – к языку гипертекстовой разметки SGML, уже использовавшемуся ранее в качестве основы разработки языка HTML. Однако они пошли иным путем – путем создания метаязыка, позволяющего пользователю определять нужные наборы тегов разметки, а не определяя в соответствии с правилами SGML некоторый конкретный набор тегов, как это было сделано при создании языка HTML.

Радикально новой целью создания платформы XML является обеспечение стандарта представления метаданных, описывающих содержание поддерживаемых в Веб информационных ресурсов. Принятый многоуровневый подход позволяет определять содержание ресурсов на нескольких уровнях – от структурных свойств типов с разной степенью глубины (XML DTD, XML Schema и Relax NG) до семантических их свойств (RDF) в терминах определенных с различной степенью формализованности понятий предметной области и их взаимосвязей (RDFS, OWL).

Здесь следует также заметить, что поддержка метаданных средствами платформы XML – отнюдь не новое слово в информационных системах. Еще во второй половине 60-х годов, как уже отмечалось выше, предложенный CODASYL принцип поддержки отчужденной от базы данных *схемы базы данных как самостоятельного ресурса* стал неотъемлемым принципом управления данными в базах данных. Этот принцип получил развитие в концепции словаря/справочника данных, предложенной в 70-х годах. Метаданные, описывающие содержание информационных ресурсов, изначально поддерживались в информационно-поисковых системах, начиная с ранних дескрипторных систем. Другие направления исследований, которые косвенным образом или непосредственно оказали влияние на развитие рассматриваемых функций Веб – это языки представления знаний в области искусственного интеллекта; семантические модели данных в технологиях баз данных, системы баз знаний, средства описания и поддержки онтологий, в особенности проекты Ontolingua (Стенфорд), DAML (Мэриленд), OIL (при поддержке Европейского союза) и др.

Таким образом, главная задача разработчиков новой технологической платформы Веб состоит в формировании сбалансированного, эффективного и эстетичного инструментария на основе использования и развития идей многих испытанных временем подходов.

4. Организация и функциональные возможности платформы XML

В отличие от действующей версии Веб, в которой все основные функции представления информационных ресурсов системы базируются на едином языке HTML, платформа XML строится на иных принципах. Разработаны "фундаментальные" стандарты (XML, XML Information Set, Namespaces in XML), составляющие концептуальную и синтаксическую основу платформы. Главным из них является стандарт XML, обеспечивающий представление основных структурных единиц информационных ресурсов нового Веб - XML-документов. Стандарт Namespaces in XML определяет для заданного XML-документа или множества документов допустимые теги разметки и их атрибуты, ассоциируя с ними по умолчанию некоторую семантику. Зарезервированные консорциумом W3C *пространства имен* используются в синтаксисе языка XML и других стандартов платформы. Namespaces in XML предоставляет, таким образом, механизм *расширения функциональности платформы*. Стандарт XML Information Set определяет набор абстрактных информационных элементов (Information Item), используемых в качестве «строительных блоков» для создания правильно построенных XML-документов.

Средствами фундаментальных стандартов определяется модульный комплекс других стандартов, определяющих языковые средства, которые обогащают функциональность языка XML дополнительными возможностями. Все эти языки используют единый синтаксис, определяемый языком XML. Именно такая модульность организации платформы XML обеспечивает ее открытый характер, возможности введения новых стандартов, не затрагивая уже существующих. Полная функциональность платформы XML определяется в настоящее время несколькими десятками взаимосвязанных стандартов, часть из которых уже принята W3C, другие находятся в стадии разработки. Эти стандарты могут использоваться на любой аппаратной платформе в любой операционной среде.

Нужно отметить, что совместно со стандартами платформы XML могут использоваться и некоторые другие стандарты, которые формально к этому комплексу стандартов не относятся. Мы будем далее называть их *окружением платформы XML*. Некоторые из таких стандартов разработаны консорциумом W3C, но не используют синтаксис языка XML. Другие стандарты окружения основаны на синтаксисе XML, но разработаны не W3C, а различными другими организациями и тем самым не имеют статуса стандартов консорциума W3C. Ряд стандартов окружения платформы XML получил достаточно широкое признание и применяется на практике. Стандарты окружения используются не только в разнообразных приложениях XML, но и в самих спецификациях стандартов платформы. Например, стандарт XPath, формально не являющийся стандартом XML, используется в спецификациях стандартов XPointer, XSLT, XQuery.

Функциональные возможности платформы XML и ее окружения можно охарактеризовать с помощью приведенной ниже функциональной классификации основных составляющих ее стандартов. В ней представлены не только уже принятые ключевые стандарты, но и ряд достаточно значимых проектов стандартов, над которыми активно ведется работа.

Основные стандарты платформы XML и ее окружения в соответствии с их функциями можно разбить на следующие классы:

- *Фундаментальные стандарты:* XML, Namespaces in XML, XML InfoSet;
- *Разметка содержания документов:* XML;
- *Определение локальной и глобальной гиперструктуры:* XPointer, XLink;
- *Форматирование и трансформация XML-документов:* XSL, XSLT, CSS;
- *Описание структуры XML-документов:* XML (DTD), XML Schema, Relax NG;
- *Описание семантики и онтологий ИП:* RDF, RDFS, OWL;
- *Языки запросов:* XQuery, XPath, XSLT;
- *Интерфейсы прикладного программирования:* DOM, SAX;
- *Обеспечение преемственности с Веб-1:* XHTML, XML Base;
- *Транспорт данных:* XMLP/SOAP, XForms;
- *Идентификация информационных ресурсов:* URI, URL, URN;
- *Информационная безопасность:* XML-Signature, XML Encryption/Decryption;
- *Архитектура функциональной надстройки Веб:* WSDL, XMLP/SOAP;
- *Вспомогательные функции:* XInclude, XFragment, Canonical XML, XPath;
- *Стандарты вертикальной сферы:* MathML, XMI, cXML, CML, WML, GML, UBL и др.

Анализируя приведенную классификацию, нетрудно видеть, что некоторые из включенных в нее стандартов многофункциональны и, в соответствии с этим, отнесены к нескольким классификационным категориям. Мы полагаем, что названия классов позволяют составить общее представление об основных функциональных возможностях платформы XML.

Более подробные сведения о конкретных упоминаемых здесь стандартах содержатся, как уже отмечалось, в их спецификациях, свободно доступных на Веб-сайте консорциума W3C [30]. Кроме того, необходимую информацию можно найти, например, в наших работах [3-6, 8-12] и в ряде изданных на русском языке монографий и учебников по языку XML.

5. Назначение и возможности языка XML

Как можно видеть из приведенной выше классификации, стандарт XML играет двойную роль в составе платформы XML. С одной стороны, он является одним из фундаментальных стандартов платформы. В этой роли его функция состоит в обеспечении *синтаксических соглашений* для определения всех других стандартов платформы. В другой своей роли язык выполняет важную функцию конструирования основных структурных единиц информационных ресурсов Веб нового поколения, называемых XML-документами.

С синтаксической точки зрения, XML является подмножеством известного языка SGML [26], допускающим сравнительно простую реализацию. В нем предусматривается набор небольшого количества операторов, главное назначение которых состоит в определении типов элементов XML-документов – «строительных блоков XML-документов» - и описании

допустимой их структуры. Таким образом, XML, как и его прототип SGML, - это метаязык, являющийся по существу языком определения тегов разметки.

Важно отметить, что в отличие от языка HTML новый язык разметки XML определяет не *форматную разметку* XML-документов, а *разметку и структурирование их содержимого* (контента). Для форматной разметки служат языки XSL и CSS.

Предполагается, что XML-документ состоит из *элементов документа* различных типов. Элементы документов могут иметь свое содержимое, а также параметры, называемые *атрибутами элемента*. Содержимое элемента может быть пустым, может быть некоторым значением и/или совокупностью элементов одного или нескольких других типов, приведенных в произвольном или некотором определенном порядке. Атрибуты элемента могут быть обязательными или факультативными и имеют скалярные значения.

Содержимое каждого XML-документа представляет собой *иерархию элементов*. При этом корневым элементом является сам документ. Каждый определяемый тип элементов документов идентифицируется некоторым именем, которое явно включается в экземпляры элементов этого типа. Имя типа элементов служит тегом разметки.

Поскольку XML определяет составные части XML-документов и допустимую их структуру, он выполняет функции *языка описания XML-данных*. Определение структуры XML-документов – типов составляющих их элементов и допустимой их структуры, представляющей иерархию таких элементов, - описывается средствами языка XML и называется *определением типа документов* (Document Type Definition). Это определение может быть встроено в конкретный документ либо хранится вне его. В последнем случае оно может определять множество экземпляров документов данного типа. В конкретном экземпляре документа для дополнения внешнего DTD может использоваться какое-либо внутреннее DTD.

Фактически DTD представляет собой сконструированное с помощью XML как метаязыка определение конкретного языка разметки с собственным набором тегов и правилами конструирования XML-документов, допустимых в этом языке. Такого рода языки принято называть *словарями XML*.

Первоначально XML разрабатывался как язык разметки информационных ресурсов для нового поколения Веб. Однако он быстро нашел применение во многих других областях информационных технологий и в определяющих их стандартах. В частности, он стал играть роль языка-посредника, используемого для обмена информационными ресурсами между различного рода системами, например, для обмена метаданными между инструментами CASE, поддерживающими язык UML.

Авторы языка XML считают принципиально важным его свойством *расширяемость*, обеспечивающую новые функциональные возможности среды Веб. Это свойство языка указывается в его названии, и оно заслуживает более точной интерпретации.

Достижение расширяемости XML основано на двух факторах. Прежде всего, XML представляет собой, как уже отмечалось, язык метауровня, а не конкретный язык, состоящий из набора тегов для разметки, подобный HTML. XML – это язык описания данных. Используя его синтаксис, можно определять различные наборы тегов для разметки содержания конкретных XML-документов или коллекций структурно однородных документов.

Второй фактор – это использование пространств имен - именованных множеств символов, используемых в качестве имен типов элементов XML-документов (тегов разметки) и их атрибутов. Пространство имен позволяет явным или неявным образом ассоциировать нужную семантику с определяемыми тегами, их атрибутами и допустимыми значениями атрибутов.

Важно подчеркнуть, что рассмотренные принципы обеспечивают также расширяемость функциональных возможностей всей платформы XML. Однако для введения каждого расширения необходимо достижение консенсуса в рамках консорциума W3C. Основу каждого дополняющего XML стандарта платформы составляет некоторый набор новых тегов с их атрибутами или только атрибутов уже введенных ранее тегов, синтаксис которых может быть определен средствами языка XML и которые поддерживают требуемые новые функциональные возможности. Для каждого нового стандарта платформы XML должно быть определено пространство имен с зарезервированным именем, описывающее символы новых тегов и/или

атрибутов тегов. Синтаксис и семантика новых тегов, атрибутов тегов и значений, которые они могут принимать, определяются в спецификации нового стандарта.

Теперь можно уточнить, как следует понимать расширяемость языка XML. Не следует буквально понимать полное название языка XML – Extensible Markup Language (расширяемый язык разметки). Строго говоря, сам язык XML не является расширяемым. Пользователь не может добавить в одобренные консорциумом W3C его спецификации никаких новых конструкций. Расширяемость языка XML понимается авторами языка в том смысле, что пользователь имеет возможность вводить и пополнять определяемое средствами языка множество тегов разметки по своему усмотрению. Для этого он использует рассмотренные выше возможности языка.

6. Особенности XML-данных и их моделирование

На развитие технологий XML весьма существенное влияние оказывают концепции и подходы, сложившиеся в области баз данных. В спецификациях стандартов платформы XML приобрели право гражданства такие термины, как модель данных, ограничение целостности, схема, язык запросов. Используются концепции иерархии абстракций данных (или иначе - многоуровневого представления данных), самоописываемости данных, схемы данных, отчужденной (автономной) от самих данных. Однако природа XML-данных и подходы к их моделированию имеют свои особенности.

XML-данные. Основной структурной единицей XML-данных является *XML-документ*, логическое представление которого, как указывалось выше, - это иерархия элементов документа. По своей структуре XML-документ напоминает записи базы данных CODASYL или записи файлов базы данных весьма популярной ранее СУБД ADABAS, разработанной в 70-х годах компанией Software AG для мейнфреймов.

На практике встречаются две категории XML-документов. Документы первой категории служат для представления некоторых данных средствами синтаксиса XML таким образом, что организация самих представляющих их XML-документов в целом, исходный их вид при загрузке в базу данных не имеют существенного значения. Важно только сохранить в базе данных элементы, составляющие содержание этих XML-документов, значения этих элементов и значения их атрибутов. Такие XML-документы называют *ориентированными на данные* (Data-Centric Documents). Для XML-документов второй категории важное значение имеет не только их содержание как множества составляющих элементов, но и исходное представление этих документов. XML-документы такого рода называются *документно-ориентированными* (Document-Centric Documents) или прозаическими документами (Prose-Oriented Documents).

Документы первой категории предназначены, главным образом, для автоматизированной обработки. В то же время, документы второй категории предназначены, как правило, для восприятия человеком. Они обычно подготавливаются с помощью какого-либо текстового редактора, а затем конвертируются в XML-представление. XML-документы, ориентированные на данные, как правило, генерируются автоматически из различных источников информационных ресурсов, в том числе, и из содержимого баз данных. В отличие от них, документно-ориентированные XML-документы обычно создаются при существенном участии человека.

Гиперссылки в XML-документах. Язык XML – не единственный инструмент конструирования XML-документов. Документ, представленный средствами языка XML, может дополняться компонентами, описываемыми средствами других языков платформы XML, - формами, электронной подписью и т.д. С помощью языков XPointer и XLink в XML-документ могут встраиваться внутренние и внешние гиперссылки. Внутренние гиперссылки позволяют определять связи между компонентами данного документа и определять тем самым его гипермедийную интраструктуру. Внешние гиперссылки связывают данный документ с другими документами или их компонентами, образуя гипермедийную макроструктуру XML-данных. Гиперссылки вместе с тем определяют навигационные пути доступа к XML-данным на заданном множестве XML-документов.

Идентификация XML-документов и фрагментов документов в среде Веб. В среде Веб XML-документы могут идентифицироваться как по месту их хранения в Интернет, так и с

помощью некоторых абстрактных имен. Предусматривается возможность использования более общего вида идентификаторов ресурсов по сравнению с хорошо известным пользователям действующего Веб универсальным указателем ресурсов (Universal Resource Locator, URL). Такой идентификатор называется *универсальным идентификатором ресурсов* (Universal Resource Identifier, URI) [18]. Привычный способ идентификации с помощью URL, а также абстрактные имена ресурсов (Universal Resource Name, URN), являются частными случаями URI. В среде Веб, таким образом, возможен прямой доступ к требуемым XML-документам по их URI, а также навигационный доступ по гиперссылкам, содержащим URI целевых XML-документов.

Для идентификации фрагментов XML-документов используются стандарты XPointer и XPath. Первый из них принадлежит платформе XML, а второй - ее окружению.

Многоуровневое представление XML-данных. Как и в области баз данных, разработчики Веб нового поколения ориентируются на использование *иерархии абстракций XML-данных* – многоуровневой информационной архитектуры: «физика-логика-семантика».

Физическое представление XML-документа – это его представление в терминах «хранимых сущностей» [25], т.е. файлов, содержащих составные части данного документа. К физическому представлению относятся также аспекты кодировки (Encoding), используемой для контента документа, идентификации его компонентов с помощью URL и т.п.

Логическое представление XML-документа – иерархическая структура, образованная из элементов документа. Логическое представление может быть описано с помощью DTD и/или языка XML Schema.

Наконец, *семантическое представление XML-документов* описывается средствами стандартов RDF, RDFS или OWL.

В отличие от принципов многоуровневого представления данных в системах баз данных, платформа XML не обеспечивает строгого разделения аспектов физического и логического представления данных. Те и другие сосуществуют в XML-документах, препятствуя реализации концепции независимости данных, имеющей важное значение в технологиях баз данных.

Структурированные и слабоструктурированные XML-данные. XML-документы позволяют представлять как слабоструктурированные, так и структурированные данные.

Действительно XML-документы некоторых типов не имеют регулярной структуры. Их структура характеризуется наличием многих факультативных элементов и факультативных атрибутов элементов. К тому же язык XML поддерживает довольно бедную систему типов значений элементов и атрибутов элементов. Содержание таких документов представляет собой *слабоструктурированные данные*. Как правило, указанным свойством обладают документно-ориентированные XML-документы. Свойства слабоструктурированных XML-данных определяются разметкой содержащих их документов. Для них могут быть также заданы DTD.

Вместе с тем, существуют типы XML-документов, характеризующихся регулярной структурой относящихся к ним коллекций экземпляров документов. Содержание таких документов представляет собой *структурированные данные*. Для таких типов XML-документов может использоваться отчужденное автономное описание – DTD и/или схема на языке XML Schema. Указанные свойства характерны для документов, ориентированных на данные.

Таким образом, стандарты платформы XML обеспечивают представление не только слабоструктурированных, но и структурированных XML-данных.

Модели данных XML. В стандартах XML активно употребляется понятие *модели данных* [13, 15], возникшее и играющее фундаментальную роль в технологиях баз данных. Однако оно трактуется здесь иначе, чем в области баз данных. В то время как в технологиях баз данных модель данных понимается как метамодель для конструирования моделей конкретных предметных областей и оперирования их элементами, т.е. как инструмент моделирования предметной области, авторы стандартов XML употребляют “старомодную” (раннюю) трактовку понятия модели данных как структуры конкретного XML-документа. Такое разночтение создает неоправданные проблемы, поскольку в настоящее время ведутся работы по созданию средств интеграции информационных ресурсов баз данных и XML.

Другая проблема заключается в том, что единой функционально полной, охватывающей как структурные, так и операционные возможности модели данных, на которой бы базировались все стандарты платформы, не существует до сих пор и похоже, что в близкое время она вряд ли сможет появиться. Никакой деятельности в этом направлении в консорциуме пока не ведется. Вопросы моделирования данных обсуждаются лишь автономно в рамках спецификаций отдельных стандартов. При этом авторы имеют в виду только структурные аспекты моделирования данных. Исключение составляет стандарт DOM, о котором говорится ниже.

В разработках программных инструментальных средств, поддерживающих стандарты XML, фактически используется несколько различных альтернативных моделей данных, которые можно было бы назвать *XML-ориентированными моделями*. Среди них модели, определяемые стандартами DOM и XPath, а также парами языков (XML, XQuery), (XML, XSLT). Первая из названных моделей (DOM) – это полнофункциональная объектная модель данных. Хотя стандарт DOM не основан на синтаксисе языка XML и поэтому относится в соответствии с нашей терминологией к ее окружению, определяемые им языковые средства часто используются в качестве API в системах управления XML-данными. Поэтому объектную модель, определяемую этим стандартом, с полным правом можно отнести к категории XML-ориентированных моделей данных. Стандарт XPath также не использует синтаксис XML и позволяет идентифицировать фрагменты XML-документа как узлы или подмножества узлов иерархической древовидной структуры, представляющей документ. Наличие возможности идентифицировать интересующие пользователя фрагменты документов позволяет использовать определяемый стандартом XPath язык как язык запросов. В остальных указанных выше моделях XML выполняет функции языка определения данных, а языки XQuery и XSLT определяют их операционные возможности. Поэтому язык запросов является фактически одним из компонентов определения модели данных. Важно отметить, что указанные языки запросов не предусматривают средств обновления данных.

Метаданные XML. Как и в технологиях баз данных, информационные ресурсы XML состоят из данных и метаданных. *Метаданные* в среде XML имеют различные назначения. Они описывают структурные свойства документов, типы содержащихся в них данных, семантику данных и т.д. Для описания структурных свойств документов и типов представленных в них данных в технологиях XML используется, как и в технологиях баз данных, отчужденное от документов описание, представляющее собой автономный информационный ресурс – DTD, схему на языке XML Schema, RDF-спецификацию и т.п. Случаи, когда в XML-документах используются встроенные спецификации описания типа документа DTD, а не ссылки на внешние DTD, по существу не нарушают этого принципа, поскольку при этом DTD четко отделены от остального контента определяемых документов.

Однако в отличие от технологий баз данных, в технологиях XML наряду с автономно существующими метаданными, отчужденными от самих данных, существуют также метаданные, встроенные в XML-документы. Встроенные метаданные представляются разметкой документов и обеспечивают их *самоописываемость*. Для извлечения значений таких метаданных необходимо сканировать содержание документа с помощью синтаксического анализатора (парсера). Таким образом, в технологиях XML сосуществуют принципы самоописываемости данных и их автономного описания. Разметка XML-документа определяет при этом структурные его компоненты и описывает некоторые простейшие свойства конкретного документа. Описание свойств типов XML-документов и содержащихся в них данных обеспечивается с помощью автономных описаний, например, DTD или схемы на языке XML Schema.

Важной особенностью управления XML-данными в Веб является отсутствие централизации, полная автономия поддержки информационных ресурсов на отдельных Веб-сайтах. В связи с этим в такой среде в отличие от систем баз данных отсутствуют механизмы централизованной поддержки метаданных. Возможно, в рамках архитектурных подходов, разрабатываемых для Веб нового поколения, указанные функции будут играть специальные Веб-сервисы. Нужно заметить, что указанные проблемы не возникают в системах баз данных нового типа, предназначенных для поддержки XML-данных. Такие системы называют *XML-ориентированными базами данных*. Их характеристики будут рассмотрены далее.

7. XML-ориентированные базы данных

Интенсивное развитие новой технологической платформы Веб, быстрое расширение сферы ее практического использования привели к интенсивному наращиванию объема разнообразных информационных ресурсов XML, поддерживаемых не только в среде Веб, но и вне ее - XML-документов, составляющих их данных и описывающих их метаданных. Для эффективного управления репозиториями таких информационных ресурсов, поддерживаемых вне Веб, естественно использовать технологии баз данных, хотя такие ресурсы обладают специфической с точки зрения традиционных баз данных природой.

Для этих целей стали применяться, прежде всего, существующие реляционные, объектные и объектно-реляционные СУБД. Проводились (и проводятся до настоящего времени) многочисленные исследования, посвященные созданию эффективных методов поддержки XML-данных средствами таких систем. Главное внимание уделяется при этом разработке методов отображения XML-данных в среду реляционных баз данных, чаще всего использовавшихся для поддержки XML-данных. Учитывая актуальность рассматриваемой области информационных технологий, ведущие поставщики коммерческих СУБД стали оснащать свои системы средствами поддержки XML-данных. В настоящее время на основе таких модернизированных систем разрабатываются многие приложения, использующие XML-данные.

Новый класс систем баз данных. Наряду с этим в области баз данных началось формирование новых технологий, учитывающих специфические свойства информационных ресурсов XML. Стали разрабатываться специализированные СУБД, основанные на стандартах платформы XML и ее окружения, которые непосредственно поддерживают XML-представление данных, а также соответствующие интерфейсы конечных пользователей и интерфейсы прикладного программирования для различных языков программирования.

Базы данных, содержащие информационные ресурсы XML, стали называться *базами данных XML*, а СУБД, обеспечивающие возможности работы с такими базами данных - *XML-ориентированными СУБД*.

Разработки XML-ориентированных СУБД начались в конце 90-х годов. К настоящему времени уже сформировалась индустрия программного обеспечения таких систем. Поставки XML-ориентированных СУБД осуществляют как ведущие разработчики традиционных серверов баз данных – компании Oracle, IBM, Microsoft и др., так и ряд вновь возникших компаний. Рынок XML-ориентированных СУБД предоставляет в настоящее время пользователям несколько десятков разнообразных СУБД, а также многие другие связанные с ними программные продукты. Свой вклад в эту область вносят и отечественные разработчики (см., например [2]). Представительный каталог существующего программного обеспечения рассматриваемого назначения поддерживает Р. Буре [22].

Классификация XML-ориентированных СУБД. Различаются две разновидности XML-ориентированных СУБД: чистые и адаптированные системы.

Чистые XML-ориентированные СУБД (Native XML DBMS) – это системы, которые создавались специально для управления информационными ресурсами XML. Они обеспечивают хранение XML-документов в полном или декомпозированном виде и обеспечивают доступ к этим документам или их фрагментам только в терминах модели данных XML. В некоторых таких системах поддерживаются гиперссылки между документами и/или их компонентами, которые используются для навигации в базе данных и доступа к XML-данным.

Первоначально такие системы были предназначены для поддержки коллекций документно-ориентированных XML-документов, для которых важно при хранении в базе данных сохранять их исходный формат, не утрачивая содержащиеся в исходном документе инструкции по обработке и комментарии, встроенные в документ сущности.

Как правило, среда хранения данных в чистых XML-ориентированных СУБД основывается на хранении XML-документов как текстов, например, в отдельных файлах. Другой подход предусматривает структурированное и декомпозированное хранение документов. В первом случае при обработке документов чаще всего необходимо использовать синтаксический анализ их содержимого. Организация хранения документов во втором случае основана на технике, используемой в традиционных СУБД. Для повышения эффективности доступа к фрагментам XML-документов обычно используется индексирование по значениям

элементов и атрибутов, а также индексирование по иерархическому ключу [14, 17] – метод, используемый ранее в системах баз данных, поддерживающих иерархическую модель данных.

В чистых XML-ориентированных системах пользовательские интерфейсы и API предусматриваются только в терминах XML. Единицей запрашиваемых данных из базы данных также являются XML-документы или их фрагменты. При этом в запросах могут использоваться характеристики порядка элементов в структуре документа. Некоторые СУБД рассматриваемого вида обеспечивают наряду с поиском и выборкой также обновление содержания документов и возможность удаления документов или их компонентов из базы данных. В этих системах обычно предусматриваются механизмы управления транзакциями. Однако производительность работы в таком режиме без использования декомпозированного хранения документов ограничена в связи с тем, что синхронизационные блокировки осуществляются в этом случае на уровне полных XML-документов.

Первой среди коммерчески чистых XML-ориентированных СУБД по всеобщему признанию является программный продукт Tamino компании Software AG. Именно компания Software AG ввела термин "Native XML DBMS" (чистые XML-ориентированные СУБД), начиная поставку этого программного продукта. В Tamino роль языка определения данных играет XML DTD, а в качестве языка запросов используется XQL [29] - один из ранних языков запросов для XML. Для управления средой хранения в Tamino используются механизмы, аналогичные использованным в раннем продукте компании – СУБД ADABAS. Индексирование данных в них основано на инвертированных списках. Такая организация среды хранения хорошо приспособлена для хранения и обработки записей иерархической структуры с повторяющимися группами с переменным числом экземпляров и с векторами данных. Именно такую структуру имеют и XML-документы.

К числу чистых XML-ориентированных СУБД относятся также 4Suite (компании Four Thought), eXtensible Information Server (eXcelon Corp.), Lore (Stanford University), ozone (ozone-db.org), Sybase ASE 12.5 (Sybase), X-Hive/DB (X-Hive Corp.) и ряд других систем.

Вторую разновидность XML-ориентированных СУБД составляют СУБД, называемые *адаптированными к XML* (XML-enabled DBMS). Это традиционные реляционные, объектно-реляционные или объектные СУБД, расширенные средствами поддержки XML-данных.

Ранние версии таких систем позволяли наряду с их традиционными возможностями поддерживать коллекции XML-документов, ориентированных на данные. В этих системах предусматриваются механизмы отображения представления XML-данных в модель данных, поддерживаемую рассматриваемой СУБД. При этом выбор способа отображения часто зависит от конкретного приложения. При этом XML-документы хранятся в декомпозированном виде. Исходный формат XML-документов обычно не сохраняется. В таких системах используется представление контента хранимых XML-документов средствами их собственной модели данных – реляционной, объектной или объектно-реляционной.

В последнее время, однако, грань между функциональными возможностями чистых и адаптированных XML-ориентированных СУБД постепенно стирается. Оснащенные дополнительными возможностями для работы с XML-данными традиционные СУБД последних версий эффективно справляются с этой задачей для обоих видов XML-документов. Они стали по существу мультимодельными системами, способными работать с различными представлениями хранящихся в них XML-данных.

К числу адаптированных XML-ориентированных СУБД можно отнести объектно-реляционные серверы баз данных DB2 (IBM), Oracle 8i, Oracle 9i (Oracle Corp.), Informix (IBM), SQL Server 2000 (Microsoft), объектную СУБД Objectivity/DB, реляционные СУБД Access 2002 и FoxPro (Microsoft) и др. Некоторые из этих программных продуктов имеют чрезвычайно развитое оснащение для поддержки XML-технологий. Так, сервер базы данных Oracle 9i включает довольно богатые инструментальные средства указанного назначения как для стадии разработки (Oracle XML Development Kit [28]), так и для стадии исполнения (Oracle XML DB [27]). Средства стадии разработки включают парсеры ряда языков платформы XML и ее окружения, разнообразные утилиты, средства поддержки интерфейсов прикладного программирования для языков Java, C, C++, PL/SQL и другие. Средства стадии исполнения не являются самостоятельными программными продуктами. Они представляют собой компоненты основного сервера баз данных, обеспечивающими расширение его функциональности.

Инструментальные средства XML в Oracle 9i обеспечивают поддержку целого ряда важных стандартов платформы XML, а именно XML, XML Schema, XSLT, SAX 2.0, DOM, XPath, SOAP, предоставляя разработчику системы базы данных выбор из множества альтернативных возможностей. Что касается способа отображения XML-документов в среду основного объектно-реляционного SQL-сервера базы данных Oracle, то можно выбрать нужную альтернативу в зависимости от ситуации. Например, документно-ориентированные XML-документы можно представлять как данные типа CLOB или BLOB. Документы, ориентированные на данные, могут декомпозироваться и представляться в виде табличных данных.

В работе [21] предпринята попытка представить целостную концепцию XML-ориентированных баз данных. Имеется русский перевод [1] ранней версии этого поддерживаемого и актуализируемого автором документа.

Функциональные возможности XML-ориентированных СУБД. Рассмотрим кратко функциональные возможности XML-ориентированных СУБД.

Прежде всего, следует отметить, что в отличие от демократичной и открытой для публикации ресурсов среды Веб, где возможно только локальное (в рамках отдельных Веб-сайтов) администрирование информационными ресурсами, для баз данных XML по необходимости обеспечивается централизованное администрирование данными.

В XML-ориентированных СУБД предусматривается более эффективная организация среды хранения информационных ресурсов XML по сравнению с используемой в Веб файловой организацией. Так, например, для доступа к XML-данным применяется техника индексирования [14, 17, 27], аналогичная используемой в традиционных технологиях баз данных.

Благодаря поддержке метаданных, описывающих свойства XML-документов, появляется возможность автоматической верификации логической целостности данных, а также решения других задач управления данными. Поиск и выборка данных в XML-ориентированной базе данных возможны с достаточно мелкой гранулярностью – с точностью до фрагментов XML-документов, обладающих требуемыми свойствами. Эту возможность обеспечивают языки запросов XQuery и XPath.

Большинство XML-ориентированных СУБД обеспечивает мультипользовательский режим работы с развитыми средствами управления транзакциями. СУБД рассматриваемой категории открывают новые возможности для разработки приложений систем баз данных.

Существующие XML-ориентированные СУБД предназначены, как правило, для управления коллекциями отдельных, независимых друг от друга XML-документов. Гиперссылки между документами и/или фрагментами документов, для определения которых служат языки XLink и XPointer, средствами большинства из этих систем не поддерживаются. Однако все-таки имеются системы, располагающие такими возможностями. Физические указатели, реализующие гиперссылки в среде хранения, позволяют существенно быстрее осуществлять поиск, формировать производные документы из взаимосвязанных компонентов.

Как и любая другая система управления базами данных, XML-ориентированная СУБД должна базироваться на модели данных, определяющей организацию и свойства данных в базе данных, а также доступные пользователям операционные возможности. Эта модель воплощается в поддерживаемых системой языковых средствах определения данных и манипулирования данными и/или в языке запросов.

К сожалению, как уже указывалось, в настоящее время не существует единой модели данных, на которой основывались бы все стандарты платформы XML. Одни из этих стандартов содержат спецификации языков, обладающих только дескриптивными возможностями, и могут тем самым использоваться как языки определения данных, описывающие те или иные их свойства. К числу таких стандартов относятся XML DTD, XML Schema, Relax NG. Некоторые другие стандарты, например язык XSLT, позволяющий описывать трансформации XML-документов, или XPath, обеспечивающий идентификацию их фрагментов, могут служить языками запросов в XML-ориентированных СУБД. В стадии разработки в консорциуме W3C находится стандарт XQuery, определяющий развитый язык запросов XML-данных. Создаваемая версия этого языка не позволяет обновлять данные. Консорциум XML:DB

разрабатывает язык запросов XUpdate, который даст возможность обновлять данные в базах данных XML.

Большинство чистых XML-ориентированных СУБД поддерживает какую-либо комбинацию дескриптивного языка и языка запросов из числа перечисленных выше.

В некоторых XML-ориентированных СУБД используются также ранее созданные языки запросов, выступавшие в качестве претендентов на роль стандарта консорциума W3C – языки XML-QL, XQL и др.

Наряду с интерактивными интерфейсами конечных пользователей для XML-ориентированных СУБД разработаны также спецификации интерфейсов прикладного программирования. В частности, для этих целей может использоваться разработанная консорциумом W3C спецификация объектной модели документов DOM (Document Object Model) [23]. На ее основе создан ряд реализаций интерфейсов прикладного программирования, обеспечивающих взаимодействие программ, написанных на объектных языках, с XML-ориентированными СУБД, поддерживающими указанный интерфейс. Другая довольно широко распространенная спецификация для рассматриваемых целей – это разработанная в рамках деятельности консорциума OASIS спецификация SAX (Simple API for XML) [16]. Первоначально SAX был предназначен для использования в Java-программах. Он быстро приобрел популярность и стал стандартом де-факто. Существующая в настоящее время версия SAX 2.0 применима в среде нескольких языков программирования.

Указанные средства позволяют оперировать представлениями XML-документов как иерархическими структурами данных (стандарты платформы XML, языки XPath и XQL, XUpdate, спецификация SAX) либо их объектными представлениями (стандарт DOM).

Значительный вклад в развитие функциональных возможностей XML-ориентированных СУБД вносит консорциум XML:DB [31]. В настоящее время в сферу его деятельности входит создание промышленных стандартов интерфейса прикладного программирования для XML-ориентированных СУБД, языка запросов для таких систем, допускающего в отличие от XQuery обновление данных в базе данных, а также создание инструментария эталонного тестирования XML-ориентированных СУБД для оценки их производительности.

Большое влияние на развитие функциональных возможностей СУБД нового класса оказывают тенденции *интеграции информационных ресурсов*, представленных в различных средах. Так, упоминавшийся выше разработанный консорциумом W3C стандарт DOM, используемый в качестве спецификации API для XML-ориентированных СУБД, обеспечивает интеграцию ресурсов баз данных XML в объектные среды. Разрабатываемый язык запросов XML (стандарт XQuery) позволяет оперировать не только XML-данными, но и реляционными данными, а также данными иерархической структуры. Адаптированные к XML объектно-реляционные СУБД, которые будут поддерживать новый стандарт SQL:2003, с помощью компонента SQL/XML смогут, по замыслу его разработчиков, обеспечить совместное использование информационных ресурсов SQL и XML. Спецификации SQL/XML позволят определять представление реляционных схем баз данных и данных SQL в форме XML-документов и наоборот представление XML-документов и метаданных XML в терминах SQL, представление операций манипулирования данными SQL (вставка, удаление, обновление) в XML, а также протоколы передачи данных при совместном использовании SQL и XML.

8. Конвергенция технологий XML и баз данных

Создание стандартов XML стимулировало процессы *конвергенции технологий Веб* и технологий баз данных. С одной стороны, сформировался новый класс систем баз данных – XML-ориентированные базы данных - и рождается индустрия специализированного программного обеспечения для таких систем. Кроме того, XML-данные поддерживаются в настоящее время большинством распространенных серверов баз данных, которые могут квалифицироваться теперь как СУБД, адаптированные к XML.

Можно утверждать, что к настоящему времени уже сложились основы нового направления технологий баз данных, призванного обеспечить эффективную технику для управления информационными ресурсами XML с учетом специфических особенностей их природы и использования.

С другой стороны, как было показано выше, в процессе развития технологии XML заимствуют важные концепции и подходы традиционных технологий баз данных - модель данных, метаданные, схема данных, язык запросов, многоуровневое представление данных, ограничение целостности данных, неопределенное значение, интерфейс прикладного программирования, транзакция и др.

Взаимное проникновение технологий XML и технологий баз данных проявляется и в разработке инструментальных средств этих технологий, осуществляемых с учетом необходимости обеспечения интеграции поддерживаемых их средствами информационных ресурсов. В качестве примера можно привести "встречное движение" разработчиков новых технологий. Как уже упоминалось, создатели технологий XML предусматривают в разрабатываемом стандарте XQuery языка запросов для XML возможности для оперирования не только данными XML, но и реляционными данными и данными иерархической структуры. В свою очередь, компонент SQL/XML создаваемого нового стандарта языка SQL, называемого теперь SQL:2003, будет определять языковые средства для интеграции данных баз данных SQL и XML-данных.

9. Заключение

Технологии XML с полным правом рассматриваются специалистами в различных областях информационных технологий как информационные технологии XXI века. Можно смело прогнозировать довольно долгую их жизнь. Основания для этого дают следующие факторы.

Прежде всего, платформа XML обеспечивает ряд новых актуальных возможностей управления данными в среде Веб, а также во многих других областях, и она открыта для дальнейшего развития. Функциональные возможности платформы разнообразны, включают ряд альтернативных средств и имеют модульную организацию. Благодаря этому различные по степени изощренности пользователи могут найти в ней функции, адекватные их потребностям. Технологии XML, первоначально ориентированные на среду Веб, нашли применение в целом ряде других областей как горизонтальной, так и вертикальной сферы.

В силу указанных причин круг специалистов, заинтересованных в технологиях XML, интенсивно расширяется. Хотя разработка и развитие ключевых стандартов платформы XML осуществляются консорциумом W3C довольно высокими темпами, можно предположить, что для достаточно полной их практической реализации потребуется еще немало времени.

Литература

1. Буре Р. XML и базы данных. Открытые системы. – 2000. - №10.
2. Веселов В., Долженков А. Опыт построения XML-СУБД. Открытые системы. - 2002. - №6.
3. Коголовский М.Р. Перспективные технологии информационных систем. - М.: ДМК Пресс, Компания АйТи, 2003. - 288 с.
4. Коголовский М.Р. Стандарты XML и электронные библиотеки. Электронные библиотеки. Институт развития информационного общества. – 2003. - Том 6. - Выпуск 2.
5. Коголовский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 800 с.
6. Коголовский М.Р. Развитие стандартов XML: новые возможности и применения. Сб. трудов Второй Всероссийской конференции “Стандарты в проектах современных информационных систем”, Москва, 27-28 марта 2002 г.
7. Коголовский М.Р. Глоссарий по стандартам платформы XML. Версия 3 (15-12-2002). http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/xmlbase/glossary_XML.
8. Коголовский М.Р. Стандарты платформы XML и базы данных (обзорный доклад). Труды Третьей Всероссийской конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”, Петрозаводск, сентябрь 2001 г.
9. Коголовский М.Р. Функциональные возможности и направления использования стандартов платформы XML. Сб. трудов Всероссийской практической конференции

“Стандарты в проектах современных информационных систем”, Москва, 2001. – М.: Открытые системы, 2001. – С. 127-133.

10. Коголовский М.Р. XML: возможности и перспективы. Часть 1. Платформа XML и составляющие ее стандарты. Директор информационной службы. Январь 2001, с. 24-28.

11. Коголовский М.Р. XML: возможности и перспективы. Часть 2. Базы данных XML, семантика XML-документов, перспективы. Директор информационной службы. Февраль 2001, с. 16-20.

12. Коголовский М.Р. XML: сферы применения. Директор информационной службы. Апрель 2001, с. 10-12.

13. Коголовский М.Р. Абстракции и модели в системах баз данных. СУБД, 4-5, 1998.

14. Новиков Б.А., Хвостиченко Б.С. Индексирование путей в слабоструктурированных данных. Труды четвертой всероссийской конференция RCDL'2002. Дубна, Россия, 2002.

15. Цикритзис Д., Лоховский Ф. Модели данных. – М.: Финансы и статистика, 1986.

16. About SAX (Simple API for XML). <http://www.saxproject.org/>.

17. Barashev D, Novikov B. Indexing XML to Support Path Expressions. Proc. of the 6th East-European Conference on Advances in Databases and Information Systems (ADBIS'2002). Bratislava, Slovenia, September 2002.

18. Berners-Lee T., Fielding R., Irvine U.C., Masinter L. Uniform Resource Identifiers (URI): General Syntax. RFC 2396. August 1998.

19. Berners-Lee T., Hendler J., and Lassila O. The Semantic Web. Scientific American. May 2001. <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.

20. Bosak J., Bray T. XML and the Second-Generation Web. Scientific American. May 1999. <http://www.sciam.com/1999/0599issue/0599bosak.html>.

21. Bourret R. XML and Databases. July, 2003. <http://www.rpbouret.com/xml/XMLAndDatabases.htm>

22. Bourret R. XML Database Products. August 13, 2003. <http://www.rpbouret.com/xml/XMLDatabaseProds.htm>

23. Document Object Model (DOM) Level 2 Specification. Version 1.0. W3C Recommendation. 13 November 2000. <http://www.w3.org/TR/2000/REC-DOM-Level-2-20001113>.

24. Eisenberg A., Melton J. SQL/XML is Making Good Progress. SIGMOD Record, Vol. 31, No. 2, June 2002.

25. Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation. 6-October-2000. <http://www.w3.org/TR/2000/REC-xml-20001006>.

26. ISO 8879:1986. Information Processing - Text and Office Systems - Standard Generalized Markup Language (SGML), 1986.

27. Oracle9i XML Database Developer's Guide - Oracle XML DB. Release 2 (9.2). October 2002. Part Number A96620-02.

28. Oracle9i XML Developer's Kit Guide - XDK. Release 2 (9.2). March 2002. Part Number A96621-01.

29. Robie J., Lapp J., Schash D. XML Query Language (XQL). The W3C Query Languages Workshop. December 3-4, 1998. Boston, Massachusetts. <http://www.w3.org/TendS/QL/QL98/pp/xql.html>.

30. XML Activity/ World Wide Web Consortium. <http://www.w3.org/xml/>.

31. XML:DB Initiative for XML Databases. <http://www.xmldb.org/>

32. World Wide Web Consortium. <http://www.w3.org/>