
ВВЕДЕНИЕ

Добро пожаловать в *XML. Базовый курс, 4-е издание* — книгу, за которую я бы многое отдал, когда начинал изучать этот язык!

Когда мы писали первое издание этой книги, XML был относительно новым языком, но уже тогда он все шире и шире использовался в приложениях, которых с каждым днем становилось все больше. К моменту работы над вторым изданием XML доказал, что он — отнюдь не минутная прихоть и не мода, и использовался для невероятно широкого диапазона приложений. К моменту выхода третьего издания было очевидно, что XML — зрелая технология, которая уже начала разделяться на отдельные области. В данном издании нам пришлось классифицировать все увеличивающееся количество спецификаций, связанных с XML, в которых либо XML используется непосредственно, либо к базовой спецификации XML добавляется некоторая функциональность.

Так что же такое XML? Это язык разметки, используемый для описания структуры данных наиболее эффективными способами. Везде, где выполняется ввод-вывод, хранение или передача данных из одного места в другое, потенциально может использоваться XML. Вероятно, наиболее известное применение XML — связанное с веб (в особенности с учетом последних разработок в этой области). Однако имеется немало других приложений, в которых XML может оказаться незаменимым, например для замены (или дополнения) традиционных баз данных или для передачи финансовой информации между двумя контрагентами. Службы новостей также используют XML для распространения синдицированной новостной информации и записей в блогах.

Книга должна помочь вам изучить все, что необходимо знать о XML: что это такое, как работает, с какими технологиями взаимодействует и как его лучше всего использовать в разных ситуациях от простой пересылки данных с использованием XML до применения XML в собственных веб-страницах. Здесь вы найдете ответы на фундаментальные вопросы.

- Что такое XML?
- Как использовать XML?
- Как работает XML?
- Для чего можно использовать XML?

Для кого предназначена эта книга

Эта книга написана для тех, кто знает, что изучение XML — стоящее дело, но не уверен в этом на все сто. Вы слышали хвалебные отзывы о XML, но сомневаетесь в том, что они справедливы. Вы, возможно, используете инструментарий, который скрывает применение XML за фасадом пользовательского приложения, но хотите знать, что именно происходит за сценой. Возможно, вы занимаетесь веб-разработкой и даже знаете азы HTML; это еще один плюс, хотя никакие из описанных знаний не являются необходимыми для изучения данной книги.

Для ее освоения вам не нужны общие знания о языках разметки. Предполагается, что вы новичок в этой области, и книга структурирована таким образом, чтобы быть полезной для начинающего и быстро превратить его в эксперта. В идеале она предназначена для новичков двух типов.

- Программисты, которые в определенной мере уже знакомы с веб-программированием или методами обмена данными. Такие читатели уже знакомы с некоторыми из рассматриваемых в книге концепций, а теперь узнают и о том, как использовать XML-технологии для улучшения решений задач, над которыми они в настоящее время работают.
- Программисты, работающие в среде программирования, но не имеющие существенных знаний или опыта в разработке веб-приложений или приложений для обмена данными. Эти читатели не только узнают о том, как использовать XML в приложениях такого рода, но и познакомятся с некоторыми новыми концепциями, которые помогут им разобраться в работе данных систем.

Как организована эта книга

Мы насколько сумели упорядочили темы данной книги от простых к сложным. Если вы еще не работали с XML, начните с введения в XML в части I, “Введение”. Отсюда вы можете перейти к конкретным интересующим вас вопросам или, если хотите, прочесть всю книгу от начала до конца. Помните, что в XML масса перекрывающихся тем и что в одних разделах использованы методы, описанные в других разделах книги.

- Мы начнем с пояснения, что такое XML и почему этот язык так важен и так распространен.
- После *почему* мы рассмотрим вопрос *как*, т.е. покажем, как создается правильно построенный XML-документ.
- Затем вашему вниманию будут предложены более сложные вопросы, связанные с созданием не просто правильно построенных, а корректных XML-

документов (вы также узнаете, что означает в данном контексте слово “корректные”).

- После того как вы познакомитесь с XML и увидите его в действии, мы перейдем к программированию и расскажем о языке программирования, который можно использовать для преобразования XML-документов из одного формата в другой.
- Рано или поздно вам придется получать XML-информацию из баз данных и сохранять ее в них. Здесь вы узнаете не только о современном положении дел в области XML и баз данных, но и о том, как запрашивать XML-информацию с применением SQL-образного синтаксиса, называемого XQuery.
- XML бесполезен, если вы не умеете писать программы для чтения данных из XML-документов и создания новых XML-документов, поэтому мы вернемся к программированию и рассмотрим несколько способов создания таких приложений.
- Программировать и использовать XML в собственном бизнесе и пересылать информацию бизнес-партнеру или публиковать ее в Интернете — это “две большие разницы”. Вы познакомитесь с XML-технологиями, которые позволят вам пересылать сообщения через Интернет, публиковать информацию и находить предоставляющие ее службы.
- Поскольку мы работаем с данными в XML-формате, было бы неплохо иметь возможность красиво выводить их на экран. И такая возможность у нас есть. Мы покажем вам XML-версию HTML, называемую XHTML. Вы также познакомитесь с технологией, которую, возможно, применяли при работе с HTML-документами, — каскадными таблицами стилей CSS. CSS позволяют добавлять визуальные стили к XML-документам. Кроме того, вы узнаете, как создавать потрясающую графику и интерактивные формы с помощью XML.
- Закончится книга практическим примером, демонстрирующим применение XML в реальной ситуации. Его идеи вы сможете использовать в собственных приложениях.

О чем эта книга

Эта книга вобрала в себя все лучшее из предыдущих изданий; в нее также добавлен новый материал, отражающий последние изменения в области XML, в частности — XQuery, RSS и Atom, а также AJAX. Внесенные обновления отражают последние версии спецификаций. Кроме того, каждая глава снабжена упражнениями для проверки понимания прочитанного материала. Возможные решения поставленных задач приведены в приложении А, “Ответы к упражнениям”.

Часть I. Введение

“Введение” — это часть, с которой необходимо начинать чтение книги большинству читателей. В ней описаны назначение XML и некоторые правила построения XML-документов. Изучив эту часть, вы должны уметь читать и создавать собственные XML-документы.

Глава 1. Что такое XML

Здесь рассматриваются некоторые базовые концепции и тот факт, что XML — это язык разметки (немного напоминающий HTML), в котором можно определять собственные элементы, дескрипторы и атрибуты (известные как *словарь*). Из этой главы вы узнаете, что дескрипторы не несут информации о представлении и предназначены только для описания данных.

Глава 2. Корректный XML

В дополнение к пояснению, что такое правильно построенный XML, в этой главе рассматриваются правила (рекомендации XML 1.0 и 1.1) именования и структурирования элементов; этим правилам необходимо следовать для получения правильно построенного XML.

Глава 3. Пространства имен XML

Поскольку вы можете сами создавать дескрипторы, необходимо уметь избегать конфликтов имен при совместном использовании документов. Пространства имен предоставляют способ уникальной идентификации группы дескрипторов с применением URI. В этой главе вы научитесь использовать пространства имен.

Часть II. Проверка

Кроме правил построения корректных XML-документов, о которых шла речь в части I, “Введение”, вы, вероятно, захотите узнать, как создаются и используются различные XML-словари. В этой части вы познакомитесь с DTD, XML Schema и RELAX NG — тремя языками, определяющими пользовательские словари XML. Вы также узнаете, как использовать эти определения для проверки корректности XML-документов.

Глава 4. Определения типа документа

Вы можете определить, как должен быть структурирован XML-документ, и даже указать значения по умолчанию с помощью определений типа документа (Document Type Definitions — DTD). Если XML отвечает связанному с ним DTD, он считается *корректным* (valid) XML. В этой главе рассматриваются азы применения DTD.

Глава 5. XML Schemas

XML Schema, подобно DTD, позволяет определить, как должен быть структурирован документ. В дополнение к определению структуры документа вы можете указывать отдельные типы данных значений атрибутов и содержимого элементов. Это более мощная альтернатива DTD.

Глава 6. RELAX NG

RELAX NG — третий метод, используемый для определения структуры документа. Кроме нового синтаксиса и новых возможностей, он берет лучшее из XML Schema и DTD и объединяет в себе простоту и мощь. RELAX NG имеет два синтаксиса, которые рассматриваются в этой книге.

Часть III. Обработка

В дополнение к определению и созданию XML-документов необходимо знать, как работать с документами, как извлекать из них информацию и преобразовывать ее в другие форматы. Именно простота извлечения и преобразования информации в другие форматы делают XML таким мощным средством.

Глава 7. XPath

Язык XPath используется для определения местоположения разделов и данных в XML-документе и очень важен для множества других XML-технологий.

Глава 8. XSLT

XML может быть преобразован в другие XML-документы, HTML и иные форматы с помощью таблиц стилей XSLT, рассматриваемых в этой главе.

Часть IV. Базы данных

Создавать и обрабатывать XML-документы — это очень важно, но в конечном итоге их надо хранить. В этой части описываются стратегии хранения и получения XML-документов и их фрагментов из разных баз данных.

Глава 9. XQuery

Очень часто необходимо получить информацию из базы данных. Элегантный способ получения информации предлагает язык XQuery, построенный на базе XPath и XPath2.

Глава 10. XML и базы данных

XML идеален для структуризации данных, и в некоторых традиционных базах данных уже появилась в той или иной степени поддержка XML. Здесь основное внимание уделяется именно этому вопросу.

Часть V. Программирование

Однажды вам придется работать с XML-документом из вашего пользовательского приложения. В этой части рассматриваются две основные методологии работы с документами — объектная модель документа (Document Object Model — DOM) и простой API для XML (Simple API for XML — SAX).

Глава 11. Объектная модель документа XML

Программисты могут использовать разные языки программирования, работая с XML с применением объектов DOM, интерфейсов, методов и свойств, которые описаны в данной главе.

Глава 12. Simple API for XML (SAX)

Альтернативой DOM для работы с XML-данными из собственных программ является простой API для XML (Simple API for XML — SAX). Из этой главы вы узнаете, как использовать SAX (на примерах Java SAX API).

Часть VI. Связь

Отправка и получение данных с одного компьютера на другой часто представляет собой сложную задачу, которую упрощают некоторые технологии сообщения с помощью XML. В этой части рассмотрены RSS и синдикация содержимого, а также веб-службы и SOAP. В настоящее издание включена новая глава, посвященная Ajax.

Глава 13. RSS, Atom и синдикация

RSS — это активно эволюционирующая технология, которая используется для публикации синдицированных новостей и резюме веб-сайтов в Интернете. В данной главе не только рассматривается использование разных версий RSS и Atom, но и обсуждаются направления развития этой технологии. Кроме того, здесь демонстрируется создание простого приложения для чтения новостей, которое в состоянии работать с любой из существующих в настоящее время спецификаций.

Глава 14. Веб-службы

Веб-служба обеспечивает межкомпьютерные сообщения. В этой главе описывается, что такое веб-службы и как работать с вызовами удаленных процедур с применением XML-RPC и REST, а также вкратце затрагиваются другие важные темы, такие как SOAP.

Глава 15. SOAP и WSDL

Одна из наиболее популярных спецификаций, связанных с XML-веб-службами, — простой протокол доступа к объектам (Simple Object Access Protocol — SOAP). Используя SOAP, можно упаковывать информацию в XML-документы

и отправлять ее через Интернет для обработки. В этой главе рассматриваются SOAP и язык описания веб-служб (Web Services Description Language — WSDL), который используется при публикации веб-служб.

Глава 16. Ajax

Ajax позволяет использовать возможности JavaScript при работе с веб-службами и SOAP или REST. Шаблоны Ajax могут использоваться в веб-страницах для связи с веб-серверами без полного обновления страниц. Эта глава вошла в книгу только в настоящем, четвертом, издании.

Часть VII. Вывод на экран

Некоторые XML-технологии связаны с выводом данных, хранящихся в XML-документах. Одни из них связаны с веб, другие разработаны для приложений и мобильных устройств. В этой части в основном рассматриваются стратегии вывода информации и используемые для этого форматы.

Глава 17. CSS

Разработчики веб-сайтов уже давно применяют каскадные таблицы стилей (Cascading Style Sheets — CSS) вместе с HTML для упрощения внесения изменений в представление информации (при этом сами HTML-документы не затрагиваются). Это решение доступно и в XML; оно позволяет корректно отображать XML-документы в браузерах. Или, если вам требуется больший уровень гибкости ваших представлений, можете использовать XSLT для преобразования XML в HTML или XHTML с последующим применением CSS для стилизации полученных документов.

Глава 18. XHTML

XHTML представляет собой новую версию HTML, которая следует правилам XML. В этой главе рассмотрены различия между HTML и XHTML и показано, как XHTML может помочь сделать ваши сайты доступными для большего количества браузеров — от устаревших до последних версий для мобильных телефонов.

Глава 19. Масштабируемая векторная графика

Вы хотите создавать графику с помощью XML? SVG позволяет описывать графику с использованием векторных команд на базе XML. В этой главе вы познакомитесь с азами SVG и с более сложным приложением на базе SVG, которое может быть опубликовано в Интернете.

Глава 20. XForms

XForms представляют собой формы на базе XML, которые могут использоваться для разработки настольных приложений, бумажных форм и, конечно же,

XHTML-форм. В этой главе показаны как базовые, так и более интересные применения XForms.

Часть VIII. Вывод на экран

Изучив эту книгу от корки до корки, вы познакомились с массой XML-технологий. В этой части теория будет применена на практике: мы создадим приложение для решения реальных задач. Данный учебный пример вошел в книгу только в четвертом издании.

Глава 21. Калькулятор платежей

В этом учебном примере используется ряд рассмотренных в книге технологий XML. Калькулятор для расчета платежей создается с помощью веб-страницы с использованием XHTML и CSS, которая общается с веб-службой посредством AJAX и использует XML Schema для построения структур данных в .NET, а DOM — для вывода данных в SVG.

Приложения

В приложении А приведены ответы ко всем упражнениям книги. В остальных приложениях представлен справочный материал, который может быть полезным для применения знаний, приобретенных вами при изучении книги, в ваших собственных приложениях.

Вот общий список приложений.

- Приложение А. Ответы к упражнениям
- Приложение Б. Описание XPath
- Приложение В. Описание XSLT
- Приложение Г. Объектная модель документа XML
- Приложение Д. Справочник по атрибутам и элементам XML Schema
- Приложение Е. Справочник по типам данных XML Schema
- Приложение Ж. Краткий справочник по SAX 2.0.2

Что необходимо для изучения этой книги

Поскольку XML — технология текстовая, все, что надо для создания XML-документов, — это Блокнот или какой-либо эквивалентный редактор. Однако, чтобы оценить приведенные примеры в действии, потребуется Интернет-браузер, способный работать с XML-документами и сообщать об ошибках в них. Даже если его у вас нет, имеющиеся в книге копии экранов позволят вам увидеть, как все это выглядит на самом деле. Кроме того, учтите следующее.

- Если вы пользуетесь Internet Explorer, то у вас есть и реализация DOM, которая может пригодиться при работе над главами, посвященными этой теме.
- Некоторые из примеров требуют доступа к веб-серверу, такому как Microsoft IIS (или PWS) или Apache.
- Все прочие используемые инструменты бесплатны. В книге приводятся инструкции о том, как их можно раздобыть.

В части книги, посвященной проверке корректности документов, представлены инструкции по работе с Codeplot (<http://codeplot.com>). Мы используем его потому, что многие инструменты проверки требуют достаточного опыта в программировании или попросту слишком велики. Codeplot может использоваться для проверки правильности построения XML-документов, для их преобразования с использованием XSLT и для помощи в кодировании XHTML, CSS и SVG.

Языки программирования

В настоящей книге мы пытались продемонстрировать вездесущность XML. Некоторые примеры специфичны для Windows, но большинство из них включают информацию по работе на других платформах, таких как Linux. Многие примеры в этом издании были изменены, чтобы можно было работать под управлением любой операционной системы с любым веб-браузером.

Кроме того, мы пытались продемонстрировать использование XML в разных языках программирования, включая Java, JavaScript, PHP, Python, Visual Basic, ASP и C#. Поэтому шансы на то, что вы встретите пример, написанный на вашем любимом языке программирования, столь же велики, как и шансы встретить пример на языке программирования, который вы никогда не использовали. Всякий раз, впервые вводя в книгу новый язык, мы включали информацию о том, как загрузить и установить необходимый для его использования инструмент. Поскольку тема нашей книги — XML, на каком бы языке ни был написан приведенный пример, базовые концепции XML рассматривались со всей тщательностью.

Соглашения

Чтобы облегчить читателям работу с текстом, в книге использован ряд соглашений.

Попробуй

Практические примеры представляют собой упражнения, которые необходимо выполнить, следуя указаниям в тексте.

1. Они обычно состоят из ряда шагов.
2. Каждый шаг имеет свой номер.
3. Выполняйте действия со своей копией базы данных.

Как это работает

После каждого практического примера следует его детальный анализ.

Во врезках наподобие этой содержится важная информация, которую следует запомнить раз и навсегда. Эта информация имеет прямое отношение к тексту, расположенному до и после врезки.

Подсказки, указания, советы и отступления от текущего материала выделяются курсивом, как этот текст.

Что касается стилей текста, то примите к сведению следующее.

- Новые термины и важные слова *выделяются курсивом*.
- Имена файлов, URL и код в тексте набраны моноширинным шрифтом: `persistence.properties`.
- Код может быть представлен в одном из двух видов:
Новый и важный код выделяется полужирным шрифтом, в то время как для не столь важного или уже показанного ранее кода такое выделение не используется.

Исходный код

При работе над примерами можно как самостоятельно вводить необходимый код, так и воспользоваться готовыми файлами, загрузив их с сайта www.wrox.com. Находясь на сайте, просто найдите название книги (*Beginning XML, 4th Edition*) и щелкните на ссылке **Download Code** на странице книги.

Поскольку у многих книг схожие названия, можно искать книгу по ISBN: SB 978-0-470-11487-2.

Загруженный код просто распакуйте с помощью своего любимого распаковщика.

Опечатки

Никто не идеален, и в нашей книге тоже есть ошибки и опечатки. Если вы обнаружите ошибку в одной из наших книг, сообщите нам о ней — мы будем крайне вам признательны.

Чтобы найти список замеченных ошибок, посетите сайт www.wrox.com и найдите страницы, посвященные данной книге, как указано выше. Затем на странице книги щелкните на ссылке **Book Errata**. На соответствующей странице вы увидите список всех обнаруженных ошибок и опечаток.

Если “вашей” ошибки в списке нет, перейдите по адресу www.wrox.com/contact/techsupport.shtml и заполните форму для сообщения о новой ошибке. Мы проверим вашу информацию, внесем ее в список обнаруженных ошибок и учтем в последующих изданиях книги.

p2p.wrox.com

Если вы хотите пообщаться с авторами книг, присоединяйтесь к форумам P2P по адресу p2p.wrox.com. Форумы представляют собой веб-системы, которые предоставляют возможность отправлять сообщения, связанные с книгами издательства Wrox, и общаться с авторами и другими читателями. Форумы также позволяют подписаться и получать по электронной почте новые сообщения на выбранные вами темы.

По адресу <http://p2p.wrox.com> вы найдете несколько форумов, которые помогут вам не только при изучении книг издательства, но и при разработке собственных приложений. Чтобы присоединиться к форумам, выполните следующие действия.

1. Посетите сайт по адресу p2p.wrox.com и щелкните на ссылке **Register**.
2. Прочтите условия работы на форуме и щелкните на кнопке **Agree**.
3. Заполните форму и щелкните на кнопке **Submit**.
4. Вы получите электронное письмо с информацией о том, как завершить процесс регистрации.

Читать сообщения на форумах можно и без регистрации, но для написания собственных сообщений регистрация обязательна.

Чтобы получить дополнительную информацию о том, как правильно пользоваться форумами Wrox P2P, прочтите P2P FAQ. Для этого нужно просто щелкнуть на ссылке **FAQ** на любой странице P2P.

ОТ ИЗДАТЕЛЬСТВА

Вы, читатель этой книги, и есть главный ее критик. Мы ценим ваше мнение и хотим знать, что было сделано нами правильно, что можно было сделать лучше и что еще вы хотели бы увидеть изданным нами. Нам интересно услышать и любые другие замечания, которые вам хотелось бы высказать в наш адрес.

Мы ждем ваших комментариев и надеемся на них. Вы можете прислать нам бумажное или электронное письмо либо просто посетить наш Web-сервер и оставить свои замечания там. Одним словом, любым удобным для вас способом дайте нам знать, нравится ли вам эта книга, а также выскажите свое мнение о том, как сделать наши книги более интересными для вас.

Посылая письмо или сообщение, не забудьте указать название книги и ее авторов, а также ваш обратный адрес. Мы внимательно ознакомимся с вашим мнением и обязательно учтем его при отборе и подготовке к изданию последующих книг. Наши координаты:

E-mail: info@dialektika.com

WWW: <http://www.dialektika.com>

Адреса для писем из:

России: 127055, г. Москва, ул. Лесная, д. 43, стр. 1

Украины: 03150, Киев, а/я 152