

Обмен данными на основе межотраслевого стандарта XML

Курочкин Михаил Владимирович, аспирант Санкт-Петербургского государственного университета культуры и искусств.

Научный руководитель, доктор технических наук, профессор Вершинина Лилия Павловна.

В процессе построения сложных информационных систем одной из главных проблем является обмен данными между различными подсистемами. Зачастую, самая простая задача интеграции данных между системами приводит к необходимости серьезных разработок. Задача существенно облегчается, если данные определенного класса будут перемещаться между подсистемами, при условии, что в этих подсистемах будет заложена технологически реализованная возможность воспринимать извне и отдавать наружу данные в стандартном формате импорта/экспорта. Данный подход является основой для разработки метаданных и интерфейсов для обмена регулярными данными для различных унаследованных разноформатных систем и основан на технологии XML.

Данный стандарт XML является обобщенным форматом данных, в переводе с английского, XML (англ. eXtensible Markup Language — расширяемый язык разметки) — рекомендованный Консорциумом Всемирной паутины язык разметки, фактически представляющий собой свод общих синтаксических правил [3].

Целью создания XML было обеспечение совместимости при передаче структурированных данных между разными системами обработки информации, особенно при передаче таких данных через Интернет. Словари, основанные на XML (например, RDF, RSS, MathML, XHTML, SVG), сами по себе формально описаны, что позволяет программно изменять и проверять документы на основе этих словарей, не зная их семантики, то есть, не зная смыслового значения элементов [5].

Из всего вышесказанного дадим следующее определение понятия XML — это иерархическая структура, предназначенная для хранения любых данных, визуальное представление структуры может быть представлено как дерево.

По мнению автора статей на сайте компании IBM, Юч Огбуджи, стандарт XML можно разделить на три группы.

В первую группу попадают базовые стандарты, с помощью которых формируются XML-документы. Такими спецификациями принято считать рекомендации консорциума W3C: XML 1.0 и XML 1.1, Namespaces in XML 1.0 и 1.1, Canonical XML, XLink (XML Linking Language), XML Base, XML Infoset, XML Schema Part 1: Structures и XML Schema Part 2: Datatypes, XPath, XPointer, а также стандарт OASIS RELAX NG (он в конце 2003 г. был признан Международной организацией по стандартизации ISO).

Вторая группа - это стандарты, относящиеся к обработке XML-документов: XSLT, DOM, XQuery, SOAP, WSDL и некоторые другие.

Третий класс XML-спецификаций составляют XML-приложения, или словари: XHTML, Docbook, XSL-FO, SVG, VoiceXML, MathML, SMIL [4].

Целесообразно расширить данную классификацию и добавить еще один вид XML-спецификаций - отраслевые XML-форматы, предназначенные для описания какой-либо конкретной отрасли, области бизнеса, хозяйственной или иной деятельности. Разработанный стандарт позволяет существенно снизить затраты на организацию информационного взаимодействия за счет унификации обмена коммерческой информацией между различными организациями: как выступающими на рынке Интернет-коммерции, так и работающими в сфере традиционной (off-line) торговли. Использование торговыми организациями программного обеспечения, поддерживающего данный стандарт, позволяет им с минимальными усилиями и без привлечения программистов организовывать публикацию своих предложений на любых поддерживающих этот стандарт Web-витринах, а также реализовать обмен информацией между собой без специальной доработки программ. Например, при оприходовании товаров у

покупателя информация о хозяйственной операции может быть автоматически загружена из данных, полученных от продавца.

Разработчики стремились обеспечить максимальную открытость стандарта с тем, чтобы он в дальнейшем мог развиваться на основании объективных потребностей рынка и поддерживаться как можно более широким кругом производителей экономического программного обеспечения и Интернет-компаний. Для этого разработчики изначально создавали стандарт независимо от особенностей собственного программного обеспечения или структур информационных баз и исходили из общих принципов организации торговой деятельности. В то же время в стандарте учтены различные особенности работы как Интернет-компаний, так и торгующих организаций.

Доказано, что XML-представление документов очень похоже на представление в иерархических СУБД. Только вместо дерева описания данных фигурирует XML-схема. Фактически, она позволяет описать все те же самые понятия в других терминах - терминах стандарта XML [6].

Следовательно при передаче разнообразных данных между разноформатными подсистемами необходимо разработать унифицированный доступ к документам, основанный на их XML-описаниях. В настоящее время существует несколько идеологий построения форматов. Следует отметить, что для описания схемы XML-файлов уже сейчас существует с десятков форматов. Однако стандартными из них являются лишь два: DTD (старый формат, являющийся частью XML 1.0) и XML Schema (утвержден в мае 2001 года).

Стандарт XML-схема является наиболее предпочтительным, так как он предоставляет значительную часть информации, которая хранится обычно в схеме базы данных (реляционной, иерархической и т.д.), а так же в файл со схемой можно внести дополнительную информацию, что позволит специализированным программам её использовать.

Для решения задач интеграции данных необходимо разработать сценарии информационного обмена, которые будут включать в себя и использовать подмножество XML-схем, что обеспечивает с одной стороны возможность работы с файлами в едином универсальном формате стандартным XML-инструментарием, а с другой стороны упрощает разрабатываемые программы для импорта/экспорта структурированных данных в XML-формате.

В процессе обмена данными между разноформатными системами по принципу DOM, загружаемые документы XML могут поступать из другого приложения, из внешнего источника данных (база данных или файл) или из формы ввода данных. Для загрузки/выгрузки данных XML в реляционные БД и документальные БД разработаны специальные программные средства, представляющие собой набор библиотек, позволяющих осуществлять загрузку и выгрузку данных в формате XML произвольной структуры.

Использование модели SAX для обработки XML-документов в данном случае представляется более разумным при реализации загрузчика, поскольку не требует значительных ресурсов памяти. Однако, при использовании модели SAX обработка документа будет происходить последовательно. Элементы документа будут обрабатываться в том порядке, в каком они встречаются в документе.

Данное обстоятельство не позволяет обрабатывать элементы документа в произвольном порядке, что может понадобиться при загрузке, в случае, если например таблицы должны быть связаны по уникальным идентификаторам со справочниками, а данные справочника расположены в документе после основных данных [1].

Можно сделать вывод, что наличие множества отраслевых схем и XML-форматов - еще одно доказательство востребованности XML-технологий, о которой упоминалось в самом начале статьи. Но если западный опыт демонстрирует устойчивую тенденцию к расширению практики применения XML, то в России дела обстоят иначе. В России с учетом

зарубежного опыта была создана общественная организация Некоммерческое партнерство "Стандарты электронного обмена информацией", цель которой - разработка национальных стандартов. Однако более чем за три года существования ею было принято всего два стандарта.

Сложившееся отношение к разработке *отраслевых* XML-форматов должно и будет меняться. Являясь частью мирового сообщества, Россия просто не сможет пойти по другому пути развития. В соответствии с мировой практикой в ближайшем будущем следует ожидать как расширения спроса на XML-форматы, так и усиления роли организаций, специально созданных для их разработки. Свидетельством тому - определенная активность государственных органов. В качестве примера можно привести УФЭБС (Унифицированные форматы электронных банковских сообщений, Банк России), язык SMML (Securities Market Markup Language, Федеральная служба по финансовым рынкам); форматы файлов, предоставляемых органам государственной власти и местного самоуправления в соответствии с Порядком предоставления в электронном виде сведений, содержащихся в Едином государственном реестре юридических лиц, Едином государственном реестре индивидуальных предпринимателей (Федеральная налоговая служба). Получается, что интерес к XML-форматам все-таки есть, причем весьма серьезный. Поэтому сейчас самое время российским компаниям и организациям, в том числе и государственным, обратить внимание на обозначенную выше мировую тенденцию.

Литература

1. Jon Bosak, Tim Bray., XML and the Second-Generation Web, Scientific American, may, 1999.
2. М.Р. Когаловский. Стандарты платформы XML и базы данных. Институт проблем рынка РАНб, 2001.
3. Адриан де Йонг, С. Е. Слэк., Разработка расширяемого и удобного в сопровождении формата на основе XML, май, 2008.
4. Юч Огбуджи, Отраслевые XML-форматы и их перспективы, июнь, 2007.
5. А. Валиков, Технология XML, СПб.:БХВ - Санк-Петербург, 2002.
6. Дидье Мартин , XML для профессионалов, Серия "Программист - программисту". Изд. "Лори", 2001.