
Кадэй Тхей (аспирант), Б. А. Шукин (д. т. н., профессор)
Московский инженерно-физический институт (государственный университет)

БЕЗОПАСНОСТЬ ХРАНЕНИЯ XML-ДОКУМЕНТОВ

В связи с формированием архивов XML-документов возникают задачи их эффективного хранения, формирования аналитических отчетов, обеспечения их безопасности. Все это предполагает использование технологий баз данных для работы с ними.

При разработке информационных систем в последнее время стали интенсивно применяться XML-технологии.

Наиболее часто эти технологии используются для работы с данными в системах контент-менеджмента и документооборота, где требуется преобразование данных в различные форматы и приведение разнообразных данных к единому представлению. В сценарии контент-менеджмента XML обычно является базовым представлением документов, из которого автоматически генерируются дополнительные представления для визуализации, печати и редактирования.

Интеграция информационных систем, осуществляемая путем асинхронного обмена сообщениями между приложениями, также активно опирается на XML-технологии. Именно в этой сфере использования XML-технологий наиболее продвинуто решение вопросов информационной безопасности.

Интенсивное использование XML-технологий приводит к появлению огромных массивов XML-документов, которые требуют специальных средств для работы с ними. XML-документы могут сохраняться в виде архивов текстовых файлов, имеющих расширение .xml. Однако это затрудняет работу с архивом, в частности, делает их уязвимыми с точки зрения безопасности, так как хакеры в настоящее время в основном сосредоточили свои усилия на краже конфиденциальных сведений. Требуется использовать средства управления базами данных для организации этих архивов.

Различают XML-документы, ориентированные на данные, и XML-документы, ориентированные на документы, в которых используется так называемая смешанная (mixed) разметка.

XML-документы, ориентированные на данные, отражают четкую структуризацию экономических документов типа счетов и широко используются в электронной торговле, в B2B системах и т. д. Информацию, сосредоточенную в этих документах, первоначально пытались сохранить в реляционных базах данных. Однако иерархическая структура XML-документов требовала распределения информации по значительному числу таблиц, что вызывало существенные неудобства.

В последнее время фирмы — производители реляционных СУБД стали включать в свои системы поддержку работы с XML-документами [1]. В реляционной таблице XML-документ представляется столбцом типа XML, а совместное использование языков SQL и XQuery позволяет формулировать запросы к подобной базе данных.

Кроме того, появились специальные («native») XML СУБД. В базах данных, поддерживаемых этими СУБД, XML-документы хранятся в специальных файлах, при этом с каждым файлом может быть связана XML-схема, которой должны соответствовать хранящиеся там XML-документы. В настоящее время разработано несколько XML СУБД, причем в России разработана СУБД «SEDNA» [2].

Обратим внимание на XML СУБД TigerLogic [3]. В отличие от других программных продуктов, претендующих на реализацию XML СУБД и использующих программный интерфейс, основанный на DOM, TigerLogic основана на модели Pick Universal Data Model



(Pick UDM) [4] и показывает масштабируемость и превосходную обработку запросов как для малых, средних, так и для больших XML-документов, оптимально используя ресурсы памяти и центрального процессора. Эта система получила высокую оценку в материалах Aberdeen Group, которая предлагает использовать ее для поддержания «оперативных» баз данных [5].

На основе модели Pick UDM разработаны СУБД D3 (TigerLogic) UNIVERSE (IBM), UNIDATA (IBM) и многие другие. В отличие от реляционной, эта модель поддерживает многозначность, единственный тип данных — строки переменной длины и определяет атрибуты как функциональные преобразования от данных, хранящихся в базе.

В соответствии с этой моделью база данных рассматривается как совокупность файлов. Стандартно «файл базы данных» состоит из «словаря», используемого для хранения атрибутов, и «области данных» — для собственно данных, хотя и те и другие организованы абсолютно одинаково. С каждым словарем может быть связано несколько областей данных, в том числе и ни одной. Это делает модель Pick UDM более удобной по сравнению с реляционной для отображения XML-документов на структуры данных, определяемые этой моделью (Рис. 1), так как основной компонент файла — запись — имеет практически неограниченную длину.

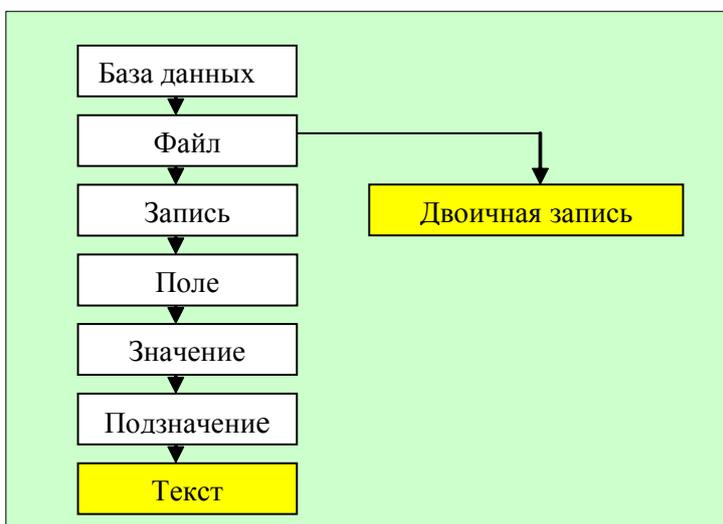


Рис. 1. Структура данных модели Pick UDM

Для отображения XML-документов будем говорить только о записях с текстовыми данными. Любая запись, будучи считанной из файла базы данных в некоторую переменную «основной памяти», например Item, принимает структуру динамического массива, к элементам которого можно обращаться по целочисленным индексам: $Item\langle i,j,k\rangle$, где i — номер поля записи, j — номер значения, k — номер подзначения. Слово «динамический» означает, что границы измерений априорно не определены и могут быть сколь угодно большими.

Модель Pick UDM не предполагает обязательную предписывающую схему, которая однозначно диктует структуру записей файлов базы данных. Так как единственным типом данных является текст, то его интерпретация может быть самой разнообразной. Поиск в базе данных может проводиться с ориентацией на произвольную описывающую схему, состоящую из атрибутов одного или нескольких словарей.

Возникает ряд задач, связанных с отображением XML-документов на структуры, определяемые моделью Pick UDM. Остановимся только на двух:

1. Конвертировать совокупность XML-документов, построенных в соответствии с некоторой XML-схемой, в файл базы данных.



2. Получить ответ на запрос к базе данных, построенной в соответствии с моделью Pick UDM, в форме XML-документа с заданной XML-схемой.

Решение связано с разработкой алгоритма отображения XML-схемы в записи определения атрибутов, сохраняемых в словаре, и с разработкой алгоритма загрузки информационной составляющей XML-документа в записи области данных.

В случае XML-документа произвольной структуры его информационная составляющая отображается в несколько записей области данных. Практически важны случаи, когда XML-документ отображается в единственную запись. Эти случаи можно описать, сформулировав следующее утверждение.

Утверждение. Для отображения информационной составляющей XML-документа в виде одной записи области данных необходимо, чтобы в дереве XML-схемы этого документа на каждом пути из корня в висячую вершину было не более двух модификаторов многозначности (символов «+» в DTD).

Большинство экономических документов удовлетворяют этому ограничению.

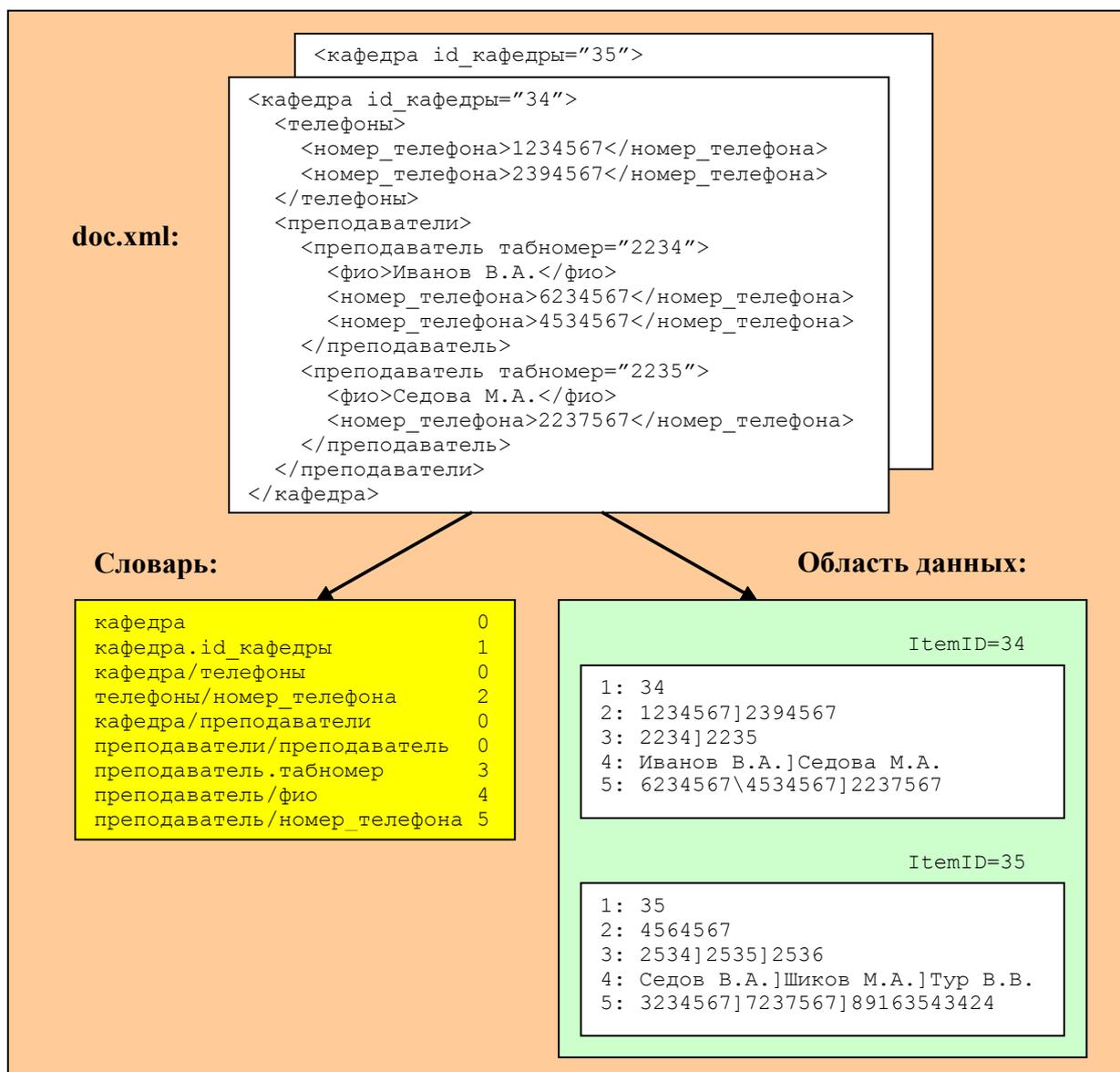


Рис. 2. Отображение XML-документов



Рассмотрим алгоритм отображения XML-схемы в структуру, определяемую моделью Pick UDM. Считаем, что XML-схема представлена в синтаксисе DTD.

Все элементы в XML-схеме необходимо расположить в стандартном иерархическом порядке в соответствии с принципом «сверху вниз, слева направо». Декларация атрибутов элемента должна следовать непосредственно за декларацией элемента, причем атрибут типа ID должен быть декларирован первым.

Рассмотрим путь от корня дерева, представляемого XML-схемой, до его висячей вершины. Примерами таких путей могут служить (Рис. 2)

кафедра/преподаватели/преподаватель.табномер

кафедра/преподаватели/преподаватель/фио

Первый путь заканчивается именем атрибута, а второй — именем негруппового элемента.

Для того чтобы иметь возможность полностью восстановить исходный XML-документ, сохраненный в базе данных в «разобранном» виде, необходимо каждому «подпути» поставить в соответствие атрибут модели Pick UDM:

кафедра

кафедра/преподаватели

преподаватели/преподаватель

преподаватель.табномер

преподаватель/фио

Таким образом, имена атрибутов модели Pick UDM образуются следующим образом:

- Корневому элементу XML-документа ставится в соответствие атрибут модели с именем корневого элемента, например: *<кафедра> -> кафедра*.

- Дочернему элементу XML-документа ставится в соответствие атрибут модели, имя которого образуется конкатенацией имени родителя, символа «/» и имени дочернего элемента:

<преподаватель> -> преподаватели/преподаватель.

- Атрибуту XML-документа ставится в соответствие атрибут модели, имя которого образуется конкатенацией имени элемента, символа «.» и имени атрибута:

<преподаватель табномер> -> преподаватель.табномер.

Формирование атрибутов модели Pick UDM производится последовательной обработкой строк XML-схемы, при этом считается, что корневой элемент всегда групповой. Всем групповым тегам ставятся в соответствие виртуальные атрибуты, которые не имеют собственного значения и которым традиционно ставится в соответствие поле с номером 0. Негрупповым тегам и атрибутам последовательно приписываются номера полей.

Загрузка XML-документа в базу данных выполняется при условии, что по XML-схеме документа предварительно построены атрибуты модели Pick UDM. Считается, что XML-документ представлен текстовым файлом с расширением .xml в каталоге windows, unix или linux. С точки зрения модели Pick UDM, это одна запись, в которой i-я строка представляет значение i-го поля. Алгоритм загрузки предполагает, что XML-документ представлен в нормализованном виде, т. е. его строки соответствуют строкам дерева элементов, получаемым при визуализации в браузере.

Предлагаемая технология позволяет создать компактную базу XML-документов, в которой средствами СУБД D3 решаются все вопросы модификации и поиска. Восстановление XML-документа в виде .xml файла или оформление ответа на запрос к базе в виде XML-документа с XML-схемой, отличной от исходной, выполняется с несущественными затратами времени. Кроме того, использование не SQL СУБД увеличивает безопасность хранения документов.



СПИСОК ЛИТЕРАТУРЫ:

1. DB2 9 pureXML Guide. URL: <http://www.redbooks.ibm.com/redbooks/pdfs/sg247315.pdf>.
2. Гринев М., Кузнецов С., Фомичев А. XML-СУБД Sedna: технические особенности и варианты использования. URL: <http://www.citforum.ru/database/articles/sedna/index.shtml>.
3. Pick Universal Data Model. URL: <http://www.infoved.ru/software/index.html#text-pick>.
4. TigerLogic XDMS. URL: <http://www.tigerlogic.com>.
5. Database Buying Guide. Fifth Edition. Aberdeen Group. October 2003. URL: www.aberdeen.com.

*В. Н. Ростовцев (к. т. н., доцент), В. А. Дамм (к. т. н., доцент),
М. В. Королев*

ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ ПОВЫШЕНИЯ НАДЕЖНОСТИ СИСТЕМ ВИБРОАКУСТИЧЕСКОГО ЗАШУМЛЕНИЯ РЕЧЕВОЙ ИНФОРМАЦИИ

Защита речевой информации от перехвата злоумышленником является важной составной частью комплексной системы защиты информации. Она достигается использованием организационных и технических методов и средств защиты. В свою очередь, технические мероприятия проводятся с использованием пассивных и активных методов. К пассивным методам относятся, например, звукопоглощение и звукоизоляция помещений. Активные методы защиты используются в том случае, когда проведением организационных мероприятий и использованием пассивных средств не обеспечивается требуемая эффективность защиты [1].

Активная защита акустической информации заключается в создании специальных шумовых помех, маскирующих речевой сигнал в среде его распространения и затрудняющих добывание этой информации противником. Активная виброакустическая маскировка эффективно используется для защиты речевой информации от утечки как по прямому акустическому, так и по виброакустическому и акустооптическому каналам утечки информации [2, 3].

Системы виброакустического зашумления (СВАЗ) позволяют обеспечить необходимый уровень защищенности ведущихся переговоров. Однако для указанных систем, как и для любых сложных технических устройств, характерно появление внезапных и постепенных отказов, при этом возникновение отказов системы зашумления может привести к неконтролируемой утечке маскируемой при помощи СВАЗ информации. Для предотвращения возможности перехвата злоумышленником защищаемой речевой информации необходимо обеспечивать надежное функционирование установленных на объектах систем виброакустического зашумления.

Для обеспечения надежной работы СВАЗ могут быть использованы различные организационно-технические мероприятия, например резервирование элементов систем зашумления. Однако не менее перспективным представляется использование и так называемого принципа обратной связи, который в данном случае можно реализовать при введении сравнительно несложных доработок в существующей аппаратуре виброакустического зашумления или при создании новых, более совершенных устройств аналогичного назначения.

