

Метод организации реляционной базы данных для слитного хранения XML документов

Альхимович А.В., Кочуров В.А., Павловский М.С.,
Цветков В.Д.

Белорусский национальный технический университет

Системы проектирования являются чрезвычайно требовательными к качеству информационного окружения конечного исполнителя - конструктора, технолога и т.д. В случае традиционного проектирования это окружение реализуется посредством доступа к различным ресурсам:

- нормативно-справочной литературе;
- знаниям более опытных коллег;
- знаниям проблемных специалистов предприятия;
- услугам различных консультационных предприятий;
- прочим источникам.

Как показывает практика, подобная схема принципиально верна, поскольку проектирование на уровне современных требований невозможно осуществить силами одного человека или небольшого замкнутого коллектива без информации извне. Эта схема доказала свою работоспособность, однако имеет и ряд недостатков:

- чрезвычайно медленный обмен данными;
- очень слабая формализация запросов, ответов и обязанностей сторон;
- субъективность оценки качества предоставляемой информации;
- уязвимость информационного пространства предприятия (особенно в условиях текучести кадров);
- сложности при взаимодействии со сторонними источниками и приёмниками информации, отсутствие стандартов подобного обмена данными (за исключением ЕСКД).

Отсюда следует необходимость создания информационной технологии, моделирующей информационное окружение проектировщика.

Ниша программных продуктов, реализующих функции доступа к справочно-эвристической информации, практически не заполнена, поскольку с технической точки зрения задача нетривиальна – необходимо создать инструментарий, который бы позволил человеку, не знакомому с информационными технологиями, воплощать свои знания в виде эвристических правил продукции.

Инженерные знания представляют собой весьма сложные, иерархические структуры данных, с огромным количеством ссылок на внешние источники информации. Кроме того, присутствует фактор вычислительных ресурсов – т.е. данные не являются пассивными (к данным могут относиться алгоритмы, формулы, правила продукции).

Однако требования сегодняшнего дня таковы, что необходимо обеспечить тесную интеграцию справочников с АРМ конструкторов, технологов, дизайнеров; ускорение доступа к справочникам, автоматическую контекстную привязку справочника в ситуации, отражённой в электронном документе, возможность автоматического обращения к внешним источникам. Так же нельзя забывать, что концепция единого информационного пространства предприятия исключает работу с бумажными документами.

Новые технологии создания распределённых информационных систем, основанные на XML Веб службах, предоставляют принципиальную возможность реализации такой задачи.

Формат представления данных XML набирает всё большую популярность. Его достоинствами являются наглядность, текстовый формат, поддержка со стороны как корпораций-гигантов (Microsoft, Sun, IBM), так и со стороны небольших фирм.

Традиционными форматами хранения XML является либо хранение в отдельных файлах, либо хранение каждого класса XML файлов в отдельной таблице. Особенность использования XML в САПР состоит в том, что здесь имеется очень много классов объектов, при небольшом количестве самих объектов. При хранении данных с использованием традиционного подхода в САПР будет образовываться большое количество файлов (таблиц), даже при объединении XML документов в один файл (таблицу). Ещё более явственно встаёт эта проблема при попыт-

ке решить задачу поиска одного и того же факта, который может встречаться в разных документах и в разных контекстах. Традиционными методами решить эту проблему не представляется возможным. Перебор всех документов неприемлем по времени выполнения операции, а реализация механизма индексирования потребует очень больших затрат ресурсов (человеческих и временных) и результатом будет низкая релевантность получаемых результатов.

В настоящей работе рассматривается слитное хранение объектов в реляционной СУБД. Слитность хранения заключается в том, что XML документ разбирается на составные части, и они последовательно сохраняются в БД. Разработанная модель БД, состоящая из фиксированного числа таблиц, позволяет хранить неограниченное число XML документов (ограничениями являются максимальный размер БД и доступное дисковое пространство).

Слитное хранение XML документов в реляционной СУБД позволит решить задачу поиска факта, который может встречаться в различных документах в различных контекстах. При традиционном подходе к решению пришлось бы последовательно открывать XML документы и осуществлять поиск средствами XML парсеров или же последовательно осуществлять поиск по всем таблицам и всем полям БД, в которой хранятся XML документы. При использовании нового подхода поиск будет осуществляться в БД с использованием индексов, что значительно увеличивает скорость поиска среди преобразованных документов. При использовании такого подхода можно осуществлять так же поиск документов, обладающих общим контекстом.

Таким образом, различные документы, спецификации, техпроцессы и т.д. могут быть представлены в виде XML документов, которые могут быть сохранены в реляционной БД в разобранном виде. Такой подход позволяет значительно ускорить поиск информации в документах за счёт использования всех возможностей современных реляционных СУБД и одновременно решает актуальную задачу поиска общего факта, который может содержаться в документах с разным контекстом.