

УДК 681.321

С.Г.Ульянов (5 курс, каф. ИУС), Д.Ф.Дробинцев, ст. преп.

РАЗРАБОТКА OLE DB-ПРОВАЙДЕРА ДЛЯ XML

В настоящее время универсальный механизм доступа к данным – Microsoft ADO (ActiveX Data Objects), широко применяемый не только в средствах разработки фирм Microsoft и Borland, но и во многих пользовательских продуктах, таких как Microsoft Office, Microsoft Internet Explorer, в ASP-приложениях и др., становится все более и более популярным способом доступа к данным, так как не только входит в состав Microsoft Office 2000 и Microsoft Internet Explorer 5.0, но и является частью ядра операционных систем семейства Windows 2000. OLE DB и ADO, являясь частью универсального механизма доступа к данным фирмы Microsoft (Microsoft Universal Data Access), позволяют осуществить доступ как к реляционным, так и к нереляционным источникам данных, таким как файловая система, данные электронной почты, многомерные хранилища данных и др.

OLE DB представляет собой программный интерфейс для доступа к различным источникам данных, таким как реляционные и нереляционные данные, текстовые, графические и географические данные, архивы электронных писем, файловая система, бизнес-объекты. В спецификации OLE DB определен набор COM-интерфейсов (COM, Component Object Model, компонентная модель объектов Microsoft, являющаяся составной частью 32-разрядных версий Windows), инкапсулирующих различные сервисы управления данными и предоставляющих однотипный доступ к перечисленным выше данным. Эти интерфейсы могут быть использованы в приложениях, предоставляющих доступ к данным, в частности, при создании провайдера данных – компонента программного обеспечения, манипулирующего данными. Он располагается между потребителем данных и базой данных. В OLE DB все провайдеры представляют данные в табличном формате (аналогичном тому, в котором хранятся данные в реляционных СУБД и файлах электронных таблиц), в виде виртуальных таблиц.

OLE DB представляет собой набор COM-интерфейсов. Отметим, что каждый OLE DB-провайдер должен содержать реализацию объектов DataSource, Session и Rowset. Помимо трех основных объектов могут быть и другие, такие как Command, Transaction и Error.

В процессе разработки подсистемы, использующей документы XML-формата, для облегчения доступа и работы с данным был реализован OLE DB провайдер данных, использующийся для доступа к XML-файлам, ядро которого стал Microsoft DOM парсер, представляющий XML документ в виде древовидной структуры. Внутри провайдера реализованы три подхода:

- при наличии только XML файла запускается эвристический разбор данного файла (по любой структуре XML пользователю предлагается несколько связанных реляционных таблиц);
- пользователь также может задать XSD-описание XML-файла. В этом случае структура таблиц будет соответствовать схемоописанию;
- пользователь также может задать XSD-описание XML-файла и директорию с однотипными XML файлами. В этом случае структура таблицы будет соответствовать схемоописанию и включать данные из всех файлов, расположенных в данном каталоге.

Данная разработка, оформленная в качестве OLE DB провайдера, позволила одинаково оперировать данными независимо от их формата (в этом и заключается основное преимущество ADO), а скорость и легкость доступа к данным оказалась на порядок быстрее существующих аналогов.