

А. Г. Израелян

Использование реляционных баз данных в обработке материалов ОЛА*

Перед составителями любого лингвистического атласа в первую очередь встает задача не создания разного рода карт, а обработки собранных ранее материалов (составление индексов), которые в дальнейшем и будут использованы для их картографической интерпретации. При работе над исходным материалом ОЛА необходимо учитывать характер задач и целей, которые будут достигнуты в результате компьютерной обработки. Главная задача проекта ОЛА – картографическое представление материалов, собранных по Вопроснику ОЛА в обследованных пунктах славянского ареала. И в этом смысле применение компьютера и различного рода баз данных значительно повышает возможности составителей ОЛА.

Первые и весьма успешные попытки использования ЭВМ в обработке данных ОЛА были предприняты Словацкой комиссией, в частности, Павлом Жиго. В 90-е годы он обратил внимание на преимущество компьютерной обработки материалов атласа и довольно подробно рассказал о своем опыте, накопленном в результате работы над материалом ОЛА, собранным им и его коллегами на территории Словацкой Республики [Жиго 1996].

Исходным пунктом использования ЭВМ при обработке материалов ОЛА было создание необходимого машинного фонда, структура которого была занесена в память компьютера в форме так называемого «поля данных». Это поле по горизонтали разделено на столбцы, в каждом из которых содержатся необходимые данные по Вопроснику ОЛА. В этих столбцах содержатся данные о номере вопроса по Вопроснику, номере пункта по сетке ОЛА, а также диалектные материалы [Жиго 1996: 57-58]. В качестве примера П. Жиго приводит ответы на вопрос № 1957 (структура машинного фонда, отражающего ответы на вопрос № 1957, § 15 Вопросника, с. 26 - таблица дана не полностью):

1957	154	sl:z, slzi:
1957	155	slzov
1957	156	selzoh
1957	208	suzi: ...

Идея создания и хранения в электронном виде материалов ОЛА, собранных на территории России, возникла давно, но по причинам различного характера начала

* Статья подготовлена при финансовой поддержке Российского гуманитарного научного фонда (проект № 07-04-00129В)

осуществляться в лишь настоящее время. Сначала было решено воспользоваться базой данных *Access*, входящей в *Microsoft Office*. *Access*-форма ОЛА (таблица) выглядит следующим образом (на примере 525 пункта):

Основная база ОЛА

Индекс	№ пункта	№ карточки	Карточка слова
...525-0015	525	0015	l'i'sa, l'i's'ica
525-0016	525	0016	l'is
525-0017	525	0017	l'i'sa, l'i's'ica
525-0018	525	0018	l'is, l'i'sa, l'i's'ica..

Но по мере увеличения объема базы данных ОЛА стало очевидно, что в дальнейшем ее возможности не будут удовлетворять требованиям, предъявляемым сотрудниками, занимающимися обработкой этих материалов. *Access* изначально рассчитан, как «настольная» база данных (возможность установки лишь на один компьютер, в который вводится и хранится информация). Его использование, как правило, ограничивается тем, что пользователь вводит в базу данных всё, что имеется в картотеке, а создание каких-либо форм не является очевидным (во всяком случае, в начале работы). Но в определенный момент стало ясно, что без аналитического подхода к решению задачи не обойтись.

Главный момент в нашей работе – это отход от интерфейсной модели, когда есть таблица, набор и запрос, к «индивидуальной» работе. В программном плане все то, что связано с составлением базы данных ОЛА, сначала было не столь очевидно, как хотелось бы, но сейчас, когда весь алгоритм действий, связанный с вводом, хранением, а главное, обработкой материалов (составление различного рода индексов) ясен, имеет смысл воспользоваться такой реляционной' (распределенной, т.е. рассчитанной на нескольких пользователей) базой данных, как, например, *Oracle Database*. Эта база данных наиболее удобна. Она является неким «фундаментом», который позволит в дальнейшем построить то «здание», которое необходимо пользователю. Кроме того, *Oracle Database* использует стандартизованный язык SQL.

SQL (Structured Query Language) – язык, разработанный специально для обслуживания реляционных баз данных. Главная цель SQL предоставить интерфейс с базой данных, например, *Oracle Database*, поэтому он может использоваться всеми: разработчиками приложений, администраторами базы данных и простыми конечными

пользователями. SQL - это неотъемлемая часть системы управления базами данных (СУБД), инструмент, с помощью которого осуществляется связь пользователя с ней.

SQL решает следующие задачи:

запросы данных;

вставка;

изменение и удаление записей в таблице;

создание, замещение, изменение и удаление объектов;

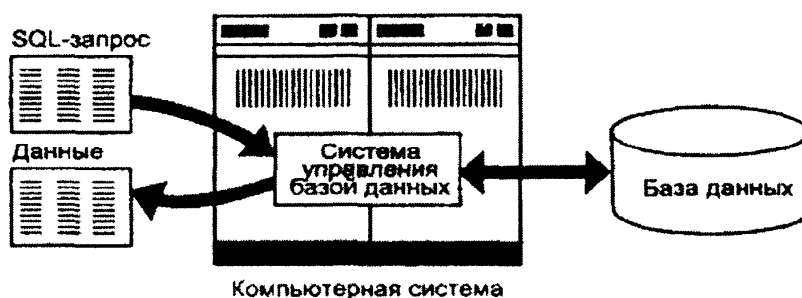
контроль над доступом к базе и ее объектам;

гарантирование устойчивости и целостности базы данных.

В отличие от традиционных языков программирования в SQL вы просто указываете то, что вам нужно и «не задумываетесь» над тем, как это сделать.

Впрочем, для лингвиста все вышеупомянутые детали технического характера не столь важны, и вряд ли стоит подробно описывать принцип работы реляционной базы данных.

Вся работа, связанная в обработкой материалов ОЛА, может быть представлена в виде простой схемы.



В нашем случае под «данными» подразумеваются различного рода индексы.

Несмотря на то, что база данных ОЛА готова пока на 35-40% (на сегодняшний день введено почти 400 тысяч карточек), очевидно, что принцип компьютерной обработки данных в высшей степени эффективен. Если при традиционном способе составления индексов (выписывание материалов из картотеки ОЛА и дальнейшая их распечатка) уходило несколько дней на составление только одного индекса, то с появлением базы данных индексы после запроса будут готовы через несколько секунд, и останется их только распечатать. Большое количество вопросов (почти три с половиной тысячи) практически не оказывает влияния на скорость работы, так как все необходимые операции компьютер по сравнению с человеком делает несоизмеримо быстрее. Более того, в

будущем, разместив готовые базы данных на сервере Института и предоставив доступ сотрудникам других национальных комиссий, можно будет лучше координировать работу различных комиссий ОЛА, готовящих очередной выпуск тома Атласа.

Очевидно, что сегодня все усилия Российской комиссии ОЛА в первую очередь должны быть направлены на то, чтобы как можно быстрее – не в ущерб качеству – ввести весь материал ОЛА (более 1 млн. карточек) в базу данных. В будущем было бы целесообразно создание **единой базы данных**, например, на основе *Oracle Database*. Правда, все вопросы о сроках создания и структуры сводной базы необходимо будет обсудить в национальных комиссиях, а главное, обеспечить их единой программой для поддержания вышеупомянутой базы данных.

ЛИТЕРАТУРА

Вопросник 1965 - Вопросник Общеславянского лингвистического атласа.

М.: 1965. Жиго 1996 - *Жиго П.* Опыт использования ЭВМ в обработке материалов ОЛА // ОЛА. Мат-лы и иссл. М., 1996.