

Проблемы создания баз данных в деятельности переводчика

*Авторы:
Ст. преподаватель каф. «НТПиПК» Пудеян Л.А.,
Студентка группы ВМПИ11 Подколзина Л.А.*

Аннотация: в статье рассмотрены основные сложности, возникающие при проектировании баз данных в области лингвистики. Предложены возможные пути решения.

Ключевые слова: лингвистика, обучение, базы данных (БД), семантическое ядро, полнотекстовые БД, фактографические БД.

В науке, и в частности в лингвистике, потребность в создании информационных систем связана с тем, что широкий доступ к необходимому исследователю информации позволяет осуществлять более результативные действия, дает возможность эффективно управлять информационными потоками. Именно поэтому актуальной задачей становится разработка междисциплинарных методик, позволяющих осуществлять аналитическую обработку информации, и поиск способов ее структурирования для эффективного хранения и обработки. Оптимальными в этой связи представляются технологии баз данных, позволяющие создавать структурированные массивы данных, хранимые и управляемые с применением компьютерных технологий, используемые для создания и функционирования эффективных информационных систем [1–5].

В современной лингвистике компьютерное представление информации на основе технологии баз данных (БД) используется в самых разных областях: в типологических и сравнительных, фонетических, лексических, грамматических исследованиях; лексикографии; семантических исследованиях, включая составление тезаурусов; в компьютерной лингвистике и для решения прикладных лингвистических задач (лингводидактических, автоматизированного перевода, автоматического распознавания и синтеза речи). В настоящий период развития этой технологии в сфере лингвистики можно говорить о двух основных типах БД:

1) полнотекстовые базы представляют собой документальные БД, в которых представлены целостные тексты, относящиеся к какой-либо дискурсивной сфере. Примером такого ресурса может выступить БД «Среднеобский фольклор» (<http://mion.tsu.ru/song>), созданная в рамках развития информационной системы «Межрегиональные исследования в общественных науках» [6];

2) собственно БД являются информационными системами фактографического типа и содержат структурированную информацию о лингвистических единицах различного рода. Например, «Хронологический морфемно-словообразовательный словарь русского языка», БД по русским прилагательным «EDGE», база фонетических данных и др. [1]

Рассмотрим основные проблемы, возникающие при создании БД. Первый этап создания любой базы данных связан с формированием системы информационных задач, на решение которых направлено создание информационного ресурса. От этой системы напрямую зависит структурирование предметной области, получающей отображение в БД, и организация ее инфологической схемы.

В случае с лингвистическими ресурсами кроме информационных потребностей следует учитывать и то, что в качестве предметной области здесь выступает фрагмент языковой системы [7], поэтому необходимо учитывать при формировании структуры уже разработанные в лингвистике описания и классификации. Еще одна проблема в отборе материала связана с активной полисемией в этой области: включать ли в БД единицы с переносным значением? От ответов на все эти вопросы непосредственно зависит структурная и функциональная специфика БД, ее информационный потенциал. Полагаем, что для реализации максимально полного информационного потенциала необходимо включение в проектируемый ресурс широкого спектра единиц, но с учетом их неравноценного положения.

Следующим этапом создания БД является проектирование дата-логической схемы БД, представляющей собой систему таблиц, поля которых отображают атрибуты описываемых объектов. В данном случае в качестве атрибутов выступают семантические, грамматические, стилистические параметры отдельных лексических единиц. Сложность параметризации семантики естественных языковых единиц отмечают практически все исследователи. Отечественные и зарубежные работы в области семантики предлагают широкий спектр методик и приемов, но полное описание плана содержания языковых единиц остается проблемой, которая не решена по сей день.

Выводы. БД является полезным и актуальным учебным материалом. Благодаря тому, что современные системы управления базами данных дополнены возможностями интеграции мультимедийных объектов. Это позволяет включить в БД данные другого типа: аудио- или видеоматериалы, демонстрирующие тип звучания, его субъект и действие, с ним связанное. Поэтому БД может быть использована в самых различных сферах преподавания языков и решения дидактических задач, что в каждом случае определяет особенности построения и функционирования БД. БД позволяет оперативно формировать комплекс КИМ, настраиваемых под конкретные условия обучения и особенности обучающихся.

Список использованных источников

1. Асиновский А.С., Архипова Е.А., Богданова Н.В. и др. Полевая лингвистическая практика: учеб.-метод. комплекс сложной структуры. Ч. 1: Теоретические основы и методика сбора лингвистических данных для представления их в речевом корпусе русского языка. СПб., 2006.
2. Гайдамакин Н.А. Автоматизированные информационные системы, базы и банки данных. Вводный курс: учеб. пособие. М.: Гелиос АРВ, 2002.
3. Герд А.С. Прикладная лингвистика. СПб.: Изд-во С.-Петербур. ун-та, 2005.
4. Советов Б.Я. Базы данных. М.: Высш. шк., 2005.
5. Хомоненко А.Д. Базы данных. СПб.: Корона-Принт, 2004.
6. Мишанкина Н.А., Тубалова И.В., Эмер Ю.А. Филология и информатика: специфика электронного представления региональных фольклорных текстов // Гуманитарная информатика: сб. ст. / под ред. Г.В. Можяевой. Томск, 2004. Вып. 1. С. 102– 114.
7. Мицкевич О.С. Лингвистическая база данных (ЛБД) специальной лексики белорусского языка с точки зрения потенциальных пользователей // Прикладная лингвистика в науке и образовании: сб. тр. VI Междунар. науч. конф., 5–7 апреля 2012 г., Санкт-Петербург. СПб., 2012. С. 203–206.