

А. А. Сурмач

*Гродненский государственный университет имени Янки Купалы
Гродно, Республика Беларусь
e-mail: surmach.anna@mail.ru*

ВЫБОР ИНСТРУМЕНТА ДЛЯ СОЗДАНИЯ СОЦИОЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ

В статье приведены требования, которым должна соответствовать социолингвистическая база данных, предназначенная для хранения и обработки результатов опросов. Описаны основные типы баз данных. Показана нецелесообразность использования для целей социолингвистических исследований реляционных баз данных и обоснована применимость документно-ориентированных баз данных.

Ключевые слова: социолингвистическая база данных; типы баз данных; требования к базам данных; социолингвистические данные; сбор социолингвистических данных.

H. Surmach

*Yanka Kupala State University of Grodno
Grodno, Republic of Belarus
e-mail: surmach.anna@mail.ru*

CHOOSING A TOOL TO CREATE A SOCIOLINGUISTIC DATABASE

The article presents the requirements that a sociolinguistic database designed to store and process results of the surveys must meet. The main types of databases are described. The inexpediency of using relational databases for the purposes of sociolinguistic research is shown and the applicability of document-oriented databases is proved.

Key words: sociolinguistic database; database types; requirements for databases; sociolinguistic data; sociolinguistic data gathering.

На сегодняшний день публикации о социолингвистическом инструментарии весьма распространены. Написано множество статей и пособий о методиках сбора социолингвистического материала. Между тем, сегодня важную роль для получения объективных результатов исследований играют большие объемы данных и их обработка. Такие данные могут быть обработаны с помощью специализированных баз данных, созданных для целей конкретных исследований. Тем не менее, следует отметить практически полное отсутствие русскоязычных публикаций, посвященных специфике создания и использования баз данных, предназначенных для сбора, хранения и обработки полевого социолингвистического материала. Такие инструменты, которые могут быть названы социолингвистическими базами данных, или СБД, имеют свою специфику.

СБД должна отвечать ряду требований, обусловленных особенностями исходных полевых данных и целями их использования. Во-первых, как и в любой базе данных, структура СБД должна обеспечивать возможность хранения всей необходимой информации. Также СБД должна иметь гибкую, адаптивную структуру. Например, если СБД создается для обработки результатов анкетирования, при изменении структуры анкеты работа с базой не должна усложняться в значительной мере. Более того, СБД должна быть масштабируема. Это значит, что СБД должна одинаково хорошо функционировать как при малом, так и большом объеме хранимых данных. СБД как минимум не должна терять своих возможностей по мере увеличения объема данных [1]. Структура СБД должна способствовать минимизации избыточности и устранять дублирование информации. Еще одним немаловажным требованием к СБД является возможность получения данных по всем необходимым запросам. Также желательно, чтобы инструмент для создания СБД был простым и понятным в использовании.

В современной технологии баз данных предполагается, что создание базы данных, ее поддержка и обеспечение доступа пользователей к ней осуществляются централизованно с помощью специального программного инструментария – системы управления базами данных (СУБД) [2]. Несмотря на то, что существует широкий спектр СУБД, условно все их многообразие может быть сведено к двум основным типам: реляционные и нереляционные СУБД. Соответственно, выделяются и два основных типа самих баз данных.

Под реляционными базами данных принято понимать «набор данных с предопределенными связями между ними», причем данные хранятся в табличной форме: «В каждом столбце таблицы хранится определенный тип данных, в каждой ячейке – значение атрибута. Каждая строка таблицы представляет собой набор связанных значений, относящихся к одному объекту или сущности. Каждая строка в таблице может быть помечена уникальным идентификатором, называемым первичным ключом, а строки из нескольких таблиц могут быть связаны с помощью внешних ключей» [3]. В качестве языка запросов при работе с реляционными базами данных используется специализированный искусственный язык SQL. С помощью SQL можно выполнять ряд операций: добавлять, обновлять и удалять строки данных, осуществлять извлечение наборов данных, а также решать иные вопросы управления базой данных. Реляционная модель баз данных является самой распространенной в мире. Разработчики программного обеспечения отмечают ряд достоинств баз данных такого типа: обеспечение целостности данных, фиксация изменений и атомарность, поддержка хранимых процедур [4]. Реляционная модель является оптимальной в тех случаях, когда данные, с которыми необходимо работать, структурирова-

ны, при этом структура не подвержена частым изменениям. Тем не менее, отмечают и недостатки этого типа баз данных: неадекватное представление сущностей реального мира, семантическая перегрузка, ограниченный набор операций, сложности при обработке рекурсивных запросов [5, с. 29].

Альтернативой баз данных реляционного типа являются нереляционные базы данных. Они устроены принципиально иначе и отличаются разнообразием. Например, документно-ориентированные базы хранят информацию в виде иерархических структур данных. Речь может идти об объектах с произвольным набором атрибутов. То, что в реляционной базе данных будет разбито на несколько взаимосвязанных таблиц, в нереляционной может храниться в виде целостной сущности. Нереляционные базы данных обладают рядом достоинств, важных для хранения и обработки социолингвистических данных. Во-первых, такая база не накладывает ограничений на типы хранимых данных. Более того, при необходимости в процессе работы можно добавлять новые типы данных [6]. Также нереляционная модель способствует быстрой разработке СБД. В последние годы в связи с резким увеличением обрабатываемых объемов данных и их сложностью началось широкое развитие нереляционных баз. Это во многом обусловлено тем обстоятельством, что, по сравнению с традиционными реляционными базами, нереляционные базы данных обеспечивают более высокую гибкость при работе в сложных средах, а также простоту модификации на самом нижнем уровне – уровне хранения данных [7, с. 739].

Необходимость разработки СБД возникла в рамках проекта, выполняемого по Государственной программе научных исследований на 2021-2025 гг. (НИР А67-21), цель которого – посредством анкетирования выявить факторы, влияющих на языковые предпочтения молодежи Гродненщины. Для того, чтобы хранить и обрабатывать материал, полученный в рамках этого исследования, в качестве СУБД была выбрана нереляционная СУБД Mongo DB. Ее выбор обусловлен несколькими причинами. Во-первых, данная документно-ориентированная база поддерживает коллекции, и все ответы респондентов могут быть разделены на такие коллекции без создания дополнительных баз данных для них. Во-вторых, документно-ориентированная база позволяет легко загружать ответы в том виде, в котором они были получены: без разделения на дополнительные таблицы и семантической перегрузки. В-третьих, работа с нереляционной базой данных не требует знания языка SQL. Несмотря на простоту его синтаксиса, данное обстоятельство может стать преградой для работы с СБД пользователей-лингвистов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. DZone: WhatDoWeMeanbyDatabaseScalability? [Electronic resource]. – Mode of access: <https://dzone.com/articles/what-do-we-mean-by-database-scalability>. – Date of access: 22.01.2022.
2. Системы управления базами данных : лекция [Электронный ресурс]. – Режим доступа: <http://www.bseu.by/it/tohod/lekcii5.htm>. – Дата доступа: 22.01.2022.
3. Реляционная база данных [Электронный ресурс]. – Режим доступа: <https://aws.amazon.com/ru/relational-database/>. – Дата доступа: 22.01.2022.
4. Что такое реляционная база данных? [Электронный ресурс]. – Режим доступа: <https://www.oracle.com/ru/database/what-is-a-relational-database/>. – Дата доступа: 22.01.2022.
5. Кузнецов, С. Д. Основы баз данных : учеб. пособие / С. Д. Кузнецов. – М. : БИНОМ. Лаборатория знаний, 2007. – 485 с.
6. SQL или NoSQL – вот в чём вопрос [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/ruvds/blog/324936/>. – Дата доступа: 22.01.2022.
7. Еремеев, А. П. Построение онтологии на основе нереляционной базы данных для интеллектуальной системы поддержки принятия решений медицинского назначения / А. П. Еремеев, С. А. Ивлиев // Программные продукты и решения. – Тверь : Изд-во ТвГТУ, 2017. – С. 739–744.