

УДК 004.65

Ивашкин Михаил Владимирович,
аспирант,
кафедры моделирования управляемых систем,
Институт экономики и управления,
ФГАОУ ВО «Уральский федеральный университет имени первого Президента
России Б.Н. Ельцина»
г. Екатеринбург, Российская Федерация

Кругликов Сергей Владимирович,
кандидат физико-математических наук, доцент, зав. кафедрой,
кафедра моделирования управляемых систем,
Институт экономики и управления,
ФГАОУ ВО «Уральский федеральный университет имени первого Президента
России Б.Н. Ельцина»
г. Екатеринбург, Российская Федерация

АНАЛИЗ ПРОТОТИПОВ И МОДЕЛИРОВАНИЕ БАЗ ДАННЫХ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Аннотация:

На основе анализа сайтов, предоставляющих официальные статистические данные, предложена модель и структура данных разрабатываемого прототипа информационного ресурса, обеспечивающего надежный доступ данных в исследовательских целях. Описан основной функционал и процессы разработки прототипа, функции которого в дальнейшем будут расширяться.

Ключевые слова:

базы данных, big data, архитектура веб-приложения.

Базы данных представляют собой структурированную и неструктурированную форму хранения информации в удобном для считывания виде. На сегодняшний день существует два основных представления хранения данных: реляционные и не реляционные. Как правило не реляционные базы данных (NoSql) зачастую связывают с технологиями BigData. В таблице 1 представлены различные сценарии применения баз данных, являющиеся не исчерпывающими. Вместе с этим они показывают возможности их применения в информационном пространстве. [1]

Таблица 1 - Сценарии использования реляционных и не реляционных баз данных

Сценарии	Реляционные базы данных	Не реляционные базы данных
1	Важна целостность данных	Требования к данным нечёткие
2	Представление данных в виде таблиц	Представление данных в виде документов «ключ-значение»
3	Более старая версия	Новая разработка
4	Более медленная скорость чтения, записи	Быстрая скорость чтения/записи

В данной работе приведен обзор официальных отечественных и зарубежных статистических ресурсов на основе которых строится алгоритм в программном и концептуальном виде, а также разрабатывается прототип создания базы статистических показателей, обеспечивающей надежный доступ к данным в исследовательских целях.

В официальной статистике применяется достаточно консервативный подход к организации баз данных и способам хранения информации. В статье [2] показано существенное отставание развития ИТ архитектуры хранения данных в области официальной статистики в сравнении с корпоративным и бизнес сегментами. Тенденция носит общий характер как для отечественных веб-ресурсов, так и для зарубежных. Наиболее используемой формой хранения официальных данных является реляционное представление, основанное на хранении данных в структурированном виде. Как правило подобные веб ресурсы имеют динамический характер предоставления информации конечному пользователю. Применяются табличное представление, позволяющее выбирать данные согласно запросам исследователя; графическое представление информации в виде графиков, схем и т.д.; расчет аналитических показателей; сохранение истории пользовательских запросов к БД на сервере и другие функции.

Менее продвинутые информационные ресурсы предлагают хранение данных в виде электронных документов, доступных для скачивания исследователю. Анализ подобных документов зачастую требует большего количества временных ресурсов для получения качественной выборки. Таблица 2 отражает обзор отечественных и зарубежных информационных ресурсов, послуживших основой для концептуальной модели базы данных [3].

Таблица 2 –Веб-ресурсы, предоставляющие статистические показатели

№	База данных	Назначение	Формат представления данных	Количество посетителей (за год), чел
Отечественные официальные статистические ресурсы				
1	Центральный банк РФ / Статистика	Основная статистика ЦБ РФ, макроэкономические показатели банковского сектора экономики РФ	DOC, XLS, PDF, ARJ	42 120 000
2	Данные Росстата	Статистические показатели уровня развития регионов	DOC, XLS, PDF, Web	11 412 000
3	JSON.TV	Исследования в области техники, интернета вещей, блокчейна, цифровизации, искусственного интеллекта и др.	DOC, XLS, PDF	3 830 000
4	ЕМИС	данные официальной статистической информации в электронном виде.	DOC, XLS, PDF, Web	1 848 000
Зарубежные официальные статистические ресурсы				
1	Statistica.com	Статистика (150 стран, 600 отраслей)	PDF,XLS,PPT	306 608 520
2	Eurostat	Данные общей региональной статистики, экономические и финансовые показатели	PNG, PDF, ZIP, TSV	262 794 522
3	World Bank Open Data	Данные экономических, социальных, финансовых показателей, природных ресурсов и окружающей среде	HTML, PDF	89 353 606
4	OECD.STAT	Данные финансовых, научно-технических, отраслевых показателей стран-участников	XLS, CSV, HTML	44 651 874

Прототип представляет собой веб-серверное приложения, направленное на реализацию базы данных в удобном для пользователя виде. База данных представлена в виде 4 таблиц. Архитектура данных и связи между таблицами представлены на рисунке 3.

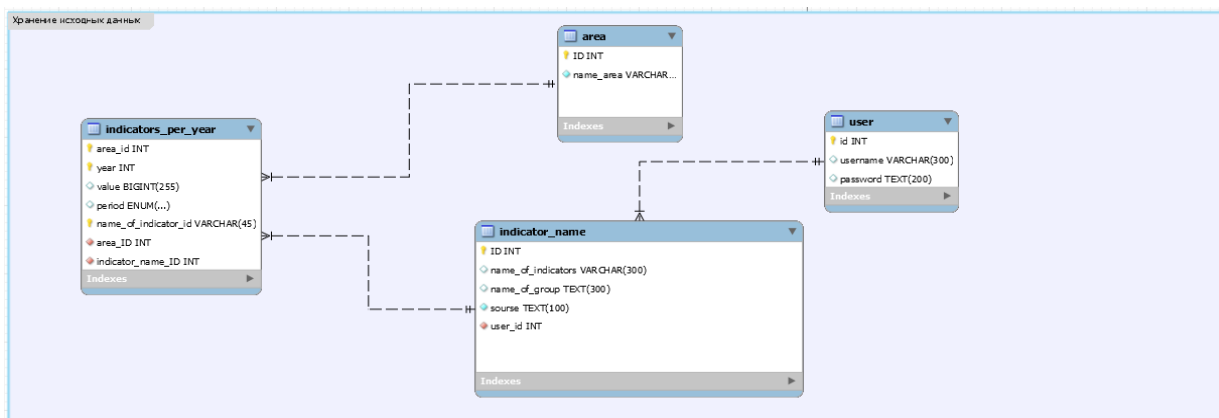
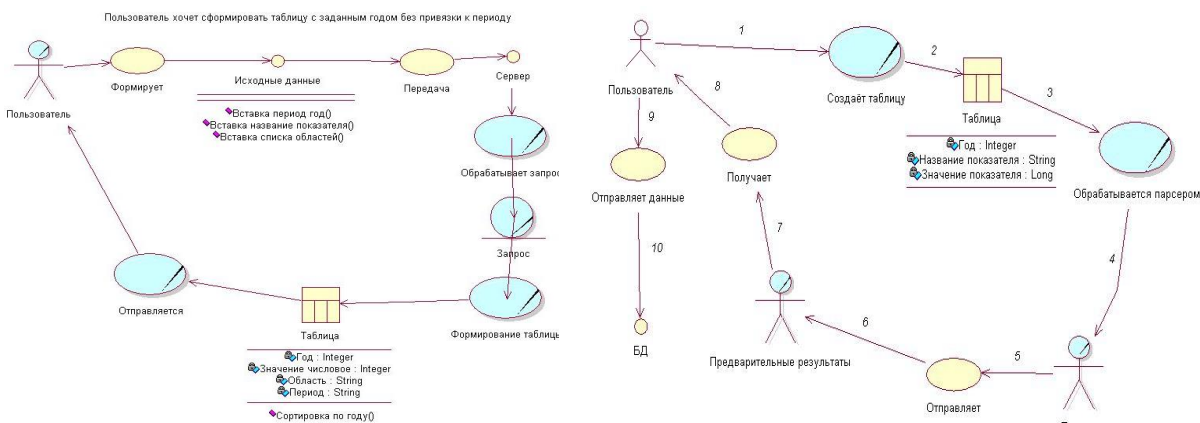


Рисунок 1 – Архитектура базы данных

Взаимодействие базы данных с веб-интерфейсом строится при помощи языка программирования Python и библиотек, обеспечивающих связку веб-интерфейса с базой данных, таких как Flask [4], Django [5] и MySQLdb [6]. Ниже, на рисунке 2 (А) представлен сценарий работы приложения, выполненный с помощью диаграммы USE-CASE [7] в программе Rational Rose. Данный сценарий выполняется каждый раз, когда пользователь желает сформировать таблицу выполнения запроса к БД. Как можно заметить – это циклический процесс, состоящий из 5 основных этапов: формирование исходных данных, передача их на сервер, в свою очередь сервер обрабатывает запрос и формирует таблицу, которая отправляется пользователю. Рисунок 2 (Б) описывает процесс загрузки данных на сервер, он реализован при помощи парсинга Excel файла, который прежде чем загрузить данные в БД, отправляется пользователю для перепроверки.

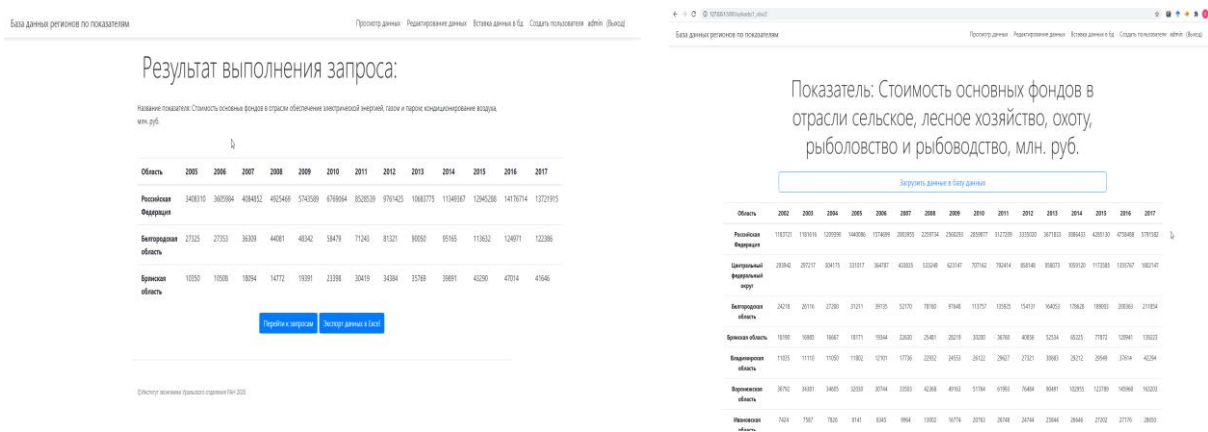


(А) Получение запроса от сервера

(Б) Процесс загрузки данных на сервер

Рисунок 2 – Взаимодействие пользователя с сервером Базы Данных

Представленные два основных сценария реализованы и успешно выполняются. На рисунке 3 представлен веб-интерфейс загрузки данных (А) и вывода результата (Б)



А
Б
Рисунок 3 – Результат выполнения запроса у пользователя

В Российской Федерации гарантом достоверности статистических данных является Федеральная служба государственной статистики. На текущий момент времени она применяет достаточно консервативное отношение к организации баз данных в официальных статистических исследованиях, однако благодаря программе «Цифровая экономика Российской Федерации» они постепенно проникают в данную область и занимают свою нишу на рынке. Текущая модель базы данных находится только на этапе проектирования и будет в будущем дополняться, и расширяться.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. <https://habr.com/ru/company/ruvds/blog/324936/>
2. Оксенойт Г. Цифровая повестка, большие данные и официальная статистика. Вопросы статистики. 2018
3. <https://changellenge.com/article/15-baz-dannykh-gde-mozhno-nayti-prakticheski-vse/>
4. <https://flask.palletsprojects.com/en/1.1.x/>
5. <https://www.djangoproject.com/start/>
6. https://mysqlclient.readthedocs.io/user_guide.html
7. ПРИМЕНЕНИЕ CASE-ТЕХНОЛОГИЙ RATIONAL ROSE И ARIS В МОДЕЛИРОВАНИИ БИЗНЕС-ПРОЦЕССОВ THE USE OF RATIONAL ROSE AND ARIS CASE-TECHNOLOGIES IN BUSINESS PROCESSES MODELLING <https://tisbi.ru/files/prod/home/nauka/vestnik-tisbi/170ee3071533126c06839b87f3cf2a43.pdf>

Ivashkin Mikhail,
Postgraduate student,
Department of Simulation of Controlled Systems,
Institute of economics and management,
Ural Federal University named after the first President of Russia B.N. Yeltsin " Yekaterinburg, Russian Federation

Sergey Kruglikov,
Candidate of physical and mathematical sciences, associate professor,
head Department of Simulation of Controlled Systems,
Institute of economics and management,
Ural Federal University named after the first President of Russia B.N. Yeltsin " Yekaterinburg, Russian Federation

ANALYSIS OF PROTOTYPES AND MODELING OF DATABASES FOR SCIENTIFIC RESEARCH

Abstract:

Based on the analysis of sites that provide official statistical data, the article describes a model according to which a prototype of an information resource is being developed, which consists of two main tasks: loading and reading data. The study also mentions the main processes by which the prototype was developed, the functions of which will be expanded in the future.

Keywords:

databases, BigData, web application architecture.