

УДК 681.3.016

Светлана Юрьевна Ржеуцкая  
Вадим Сергеевич Сафонов

## ТЕМПОРАЛЬНЫЕ БАЗЫ ДАННЫХ И ИХ ПРИМЕНЕНИЕ К АНАЛИЗУ РЕЗУЛЬТАТОВ СОЦИОЛОГИЧЕСКИХ ОПРОСОВ

В статье рассмотрено, что такое темпоральные базы данных, выделены основные перспективные направления исследований в этой области. Показан пример создания темпоральной базы данных для хранения результатов социологических опросов средствами реляционной системы управления базой данных.

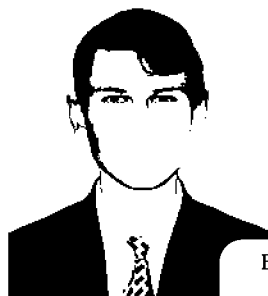
За 16 лет своего существования ВНКЦ накоплены большие объемы данных о результатах социологических опросов. Они проводятся Центром ежеквартально и охватывают несколько тысяч человек, каждая анкета включает сотни вопросов. Хорошо поставлена работа по сбору и первичной обработке результатов, анализу собранной информации, но способы ее хранения нельзя назвать оптимальными, так как результаты каждого опроса разрознены, что в некоторой степени затрудняет анализ динамики изменения данных.

В настоящее время практически все крупные информационные системы создаются с применением баз данных (БД). Теория баз данных развивается уже не одно десятилетие, но до сих пор

многие ее вопросы остаются непроработанными. Современные системы управления базами данных (СУБД) имеют богатый арсенал средств, ориентированных на работу со временем, но, к сожалению, все эти средства рассчитаны на обработку информации, содержащейся во временных рядах, а не на оперирование временными данными. Поэтому их обработка осуществляется средствами самого приложения, а не СУБД. Это порождает две взаимосвязанные проблемы. Первая из них заключается в том, что возникает необходимость разрабатывать дополнительные программные модули, что приводит к значительному усложнению информационной системы и, следовательно, к увеличению ее стоимости. Вторая проблема – это то,



С.Ю. РЖЕУЦКАЯ – к.т.н.,  
профессор ВоГТУ



В.С. САФОНОВ – аспирант  
ВНКЦ ЦЭМИ РАН

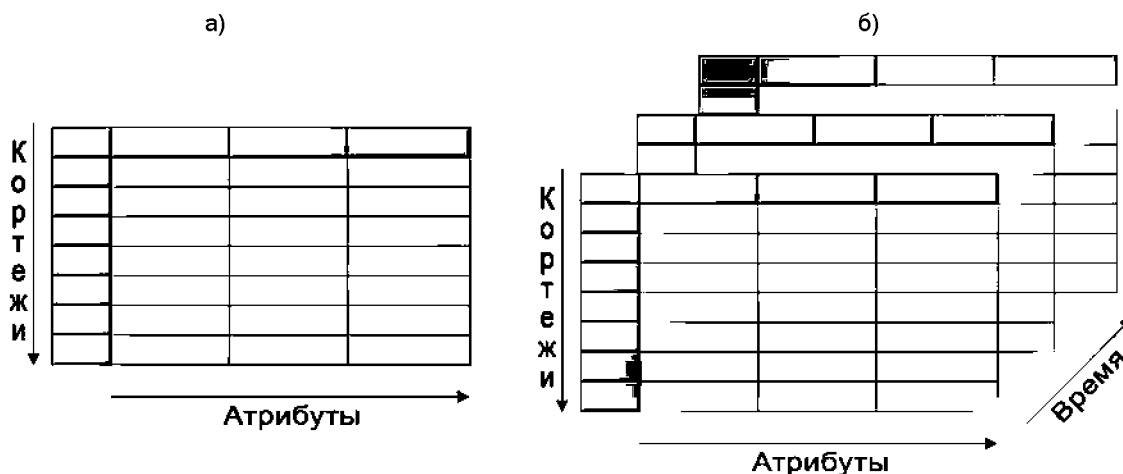
что отсутствие единого стандарта организации временной информации приводит к усложнению обмена данными между информационными системами.

В статье предложены направления исследований, подробно рассматривается возможность применения *темпоральной (привязанной ко времени) базы данных* для хранения информации, накопленной в ВНКЦ ЦЭМИ РАН в ходе проведения социологических опросов.

### Введение в темпоральные БД

В обычной реляционной БД все таблицы имеют два измерения – столбцы и строки, которые хранят мгновенный снимок предметной области. Применительно к темпоральной БД можно говорить о добавлении третьего измерения в существующие таблицы. Это позволяет хранить не только мгновенную картину, но и всю историю изменений. На *рисунке 1* схематически изображены таблицы обычной реляционной и темпоральной баз данных.

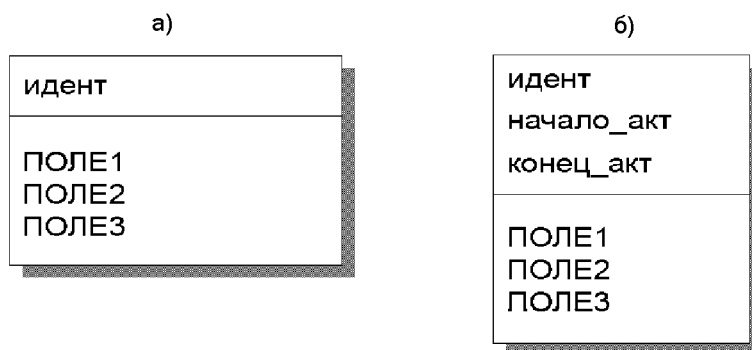
Рисунок 1. Сравнение обычных реляционных (а) и темпоральных (б) таблиц баз данных



Основное отличие обычных реляционных таблиц от темпоральных заключается в том, что они могут иметь как простые, так и составные первичные ключи, тогда как первичные ключи темпоральной таблицы всегда будут составными. Причем в состав ключа

темпоральной таблицы обязательно входят поля времени, определяющие периоды актуальности и транзакции, по отдельности или совместно. Итак, обобщенная структура темпоральных и нетемпоральных таблиц будет выглядеть, как показано на *рисунке 2*.

Рисунок 2. Структура таблиц нетемпоральной (а) и темпоральной (б) баз данных



Поддержка времени в СУБД была заложена еще в первых версиях стандарта языка SQL, когда постоянно происходило совершенствование работы с временными типами, и тем не менее можно сказать, что практика в этой области идет впереди теории.

В ходе изучения проблемы темпоральных баз данных было замечено, что их создание только еще начинается силами отдельных исследователей.

В этой проблемной области можно выделить несколько направлений исследований. *Первое* – совершенствование механизмов работы СУБД с временными данными, так как реализация одной и той же операции в различных СУБД требует написания различного объема кода, в большинстве случаев очень значительного. Это приводит к усложнению SQL-запросов и, следовательно, к увеличению времени на написание и отладку приложений, что в конечном итоге влияет на стоимость разработки. Имеет смысл создание некоторого промежуточного уровня между СУБД и клиентским приложением, который преобразовывал бы упрощенные запросы от клиентского приложения в запросы, соответствующие стандарту SQL.

При определении *второго направления* исследований можно исходить из того, что добавление временной составляющей в базу данных приводит к значительному увеличению объема хранимой информации. Это, в свою очередь, требует совершенствования индексов, содержащих временные атрибуты для ускорения обращения к временным данным.

*Третье направление* заключается в том, что в современных СУБД механизмы поддержания целостности не в полной мере соответствуют требованиям работы с темпоральными данными, поэтому необходимо совершенствование этих механизмов.

*Четвертое направление* исследований носит практический характер и состоит в разработке темпоральной базы данных для анализа на ее основе динамики данных.

Остановимся коротко на основных положениях теории темпоральных расширений реляционных БД. Прежде всего следует отметить, что выделяют четыре типа таблиц:

1. Регулярные – обычные реляционные таблицы без поддержки временных атрибутов.

2. Таблицы с временем транзакции – фиксирующие время внесения информации в БД.

3. Таблицы с временем актуальности – показывающие время, когда произошло событие в реальности.

4. Битемпоральные таблицы – хранящие оба указанных выше временных атрибута.

Наибольший интерес представляет четвертый тип таблиц. Так как в большинстве систем моменты совершения события и добавления информации в БД различаются, то имеет смысл хранить оба временных атрибута. Но, соответственно, нужно каким-то образом маркировать битемпоральные отношения, для чего существует несколько способов, сравнение которых приведено в [1].

Широко применяемые регулярные таблицы являются частным случаем битемпоральных таблиц. Таким образом, возможно преобразование любой нетемпоральной реляционной базы данных в темпоральную с сохранением всех ее свойств. И при внедрении темпоральных баз данных нет необходимости полностью уничтожить существующую базу, а достаточно лишь доработать ее, добавив поддержку времени [2].

Преобразование нетемпоральной БД в темпоральную дает следующие преимущества:

- возможность получить сведения о состоянии информации в любой момент времени в прошлом, а не только о ее текущем состоянии;
- возможность производить анализ динамики данных.

Создание единой базы социологических опросов

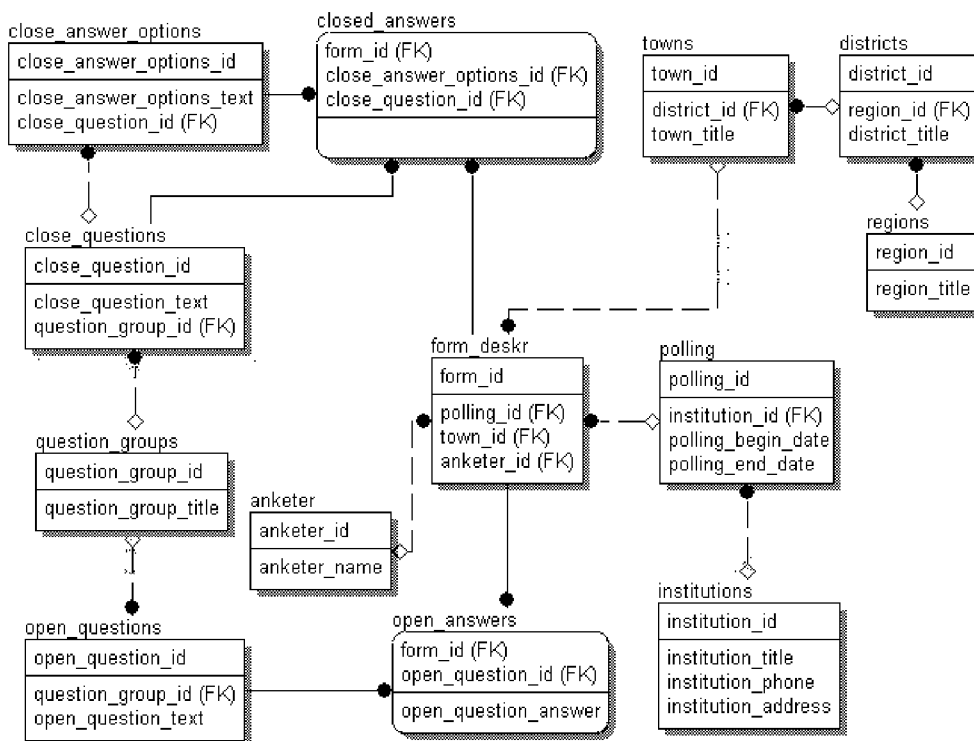
Рассмотрим задачу формирования единой БД социологических опросов. Несмотря на то, что такие данные сами по себе вернее будет отнести к информации, содержащейся во временных рядах, а не к темпоральной, некоторым аспектам этих БД добавление временных атрибутов будет полезным, поскольку с течением времени может происходить изменение вопросов или вариантов ответов. Следует обратить внимание на тот факт, что некоторые части анкет совпадают, хотя опросы могут преследовать разные цели или быть ориентированы на различные группы населения (например, различающиеся по

месту жительства или социальному положению и т. д.). Сведение данных таких опросов в единую базу позволит значительно расширить исходную информацию для анализа.

Таким образом, создание единой нормализованной реляционной базы данных является важной задачей, а добавление темпоральных атрибутов в те таблицы, где это необходимо, позволит расширить ее возможности.

Результаты за все периоды измерений сформированы в виде файлов программы SPSS и по сути дела представляют таблицу, в каждой строке которой хранятся данные одной анкеты. Сведение в единое целое данных опросов, проведенных в различные периоды, в такой форме весьма затруднительно, хотя этот вид представления удобен для анализа отдельно взятых результатов. По нашему мнению, наиболее подходящий способ хранения информации – нормализованная реляционная БД (рис. 3).

Рисунок 3. Предлагаемая логическая структура БД без учета времени



Рассмотрим данную логическую схему, из которой видно, что результаты ответов на анкеты хранятся в двух таблицах – `closed_answers` и `open_answers` (соответственно для закрытых и открытых вопросов). Таблица `closed_answers` соединена идентифицирующей связью с таблицами `form_descr` (описание анкеты), `closed_answers_options` и `closed_questions`, а таблица `open_answers` – с таблицами `open_questions` и `form_descr`. В базе предусмотрена возможность формирования групп вопросов (они хранятся в таблице `question_groups`). Таблица, хранящая информацию о каждой анкете, соединяется не идентифицирующими связями с таблицами, описывающими опрос, в рамках которого производится заполнение анкеты (`polling`), а также место заполнения анкеты и др.

Хотя сами данные опросов не являются временными данными, но из логической схемы видно, что для некоторых таблиц может быть добавлен временной аспект: например, для группы таблиц, описывающих место заполнения анкеты. Несмотря на то, что административно-территориальное деление достаточно постоянно, все же в силу сложившейся в последнее время тенденции укрупнения административно-территориальных единиц необходимо учитывать такое изменение в разрабатываемой БД. Без введения временной составляющей поддерживать целостность данных будет затруднительно.

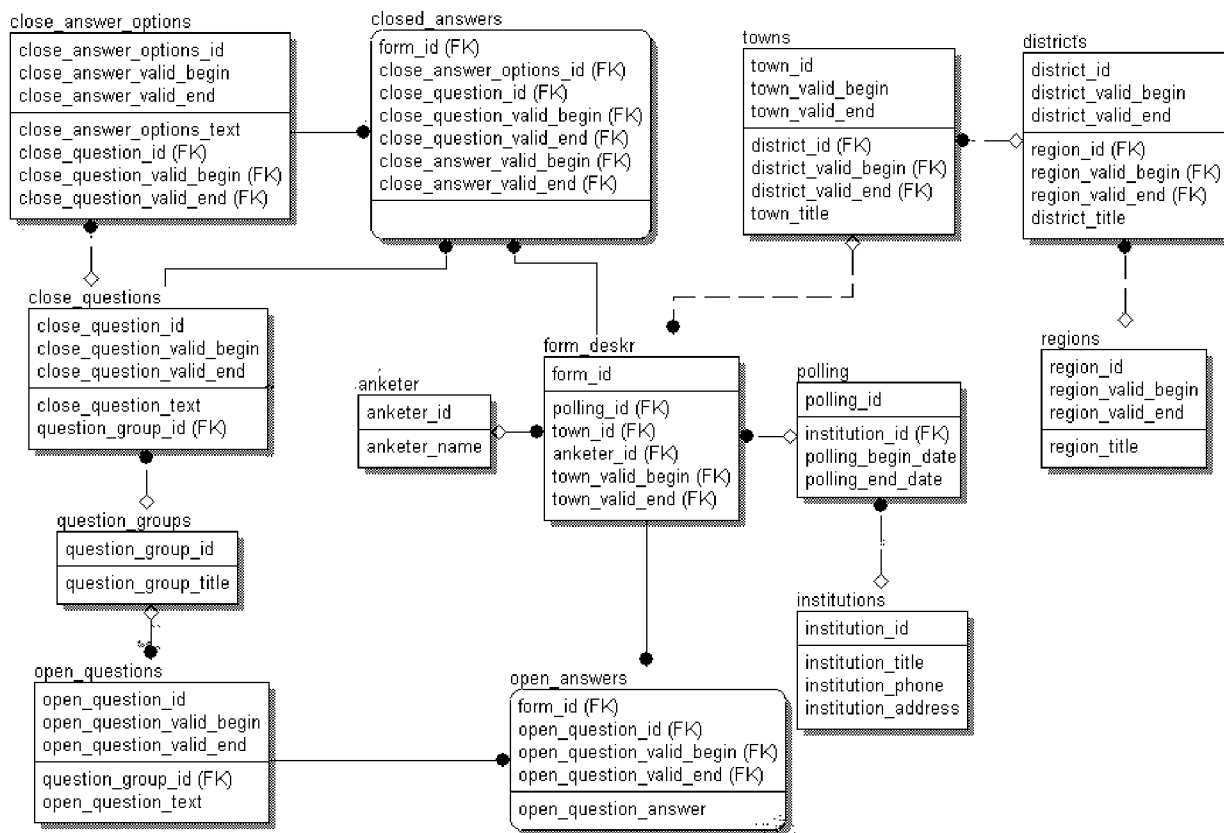
Кроме того, временной аспект необходим в таблицах, содержащих вопросы и варианты ответов на них, так как с течением времени происходит их постепенное изменение.

В представленной схеме базы данных первичные ключи всех таблиц, которые предлагается преобразовать в темпоральную форму, являются простыми и представляют собой некоторый уникальный идентификатор, удовлетворяющий нормальным формам и обеспечивающий неповторимость кортежа. При преобразовании базы в темпоральный вид применение простых первичных ключей для этих таблиц невозможно, поскольку не будет обеспечена уникальность записей, поэтому необходимо использовать составные первичные ключи вида [идентификатор записи, начало периода актуальности, конец периода актуальности]. И тогда схема базы данных будет иметь следующий вид (рис. 4).

Если сравнить рассмотренные схемы, можно заметить, что они отличаются структурой первичных ключей. Хотя преобразование в темпоральный вид расширяет возможности БД, но при этом может происходить некоторое усложнение запросов. Запрос к первой базе данных, возвращающий результаты ответов на все закрытые вопросы всех анкет всех опросов, имеет следующий вид:

```
SELECT * FROM
    form_descr, polling, anketer, closed_answers,
    close_answers_options, close_questions
WHERE
    form_descr.polling_id = polling.polling_id AND
    form_descr.anketer_id = anketer.anketer_id AND
    closed_answers.form_id = form_descr.form_id AND
    closed_answers.close_answers_options_id = close_answers_options.close_answers_options_id AND
    closed_answers.close_question_id = close_questions.close_question_id
```

Рисунок 4. Предлагаемая структура СУБД с учетом времени



Но, несмотря на одинаковый запрос, информация для второй базы будет более подробна и достоверна. Так, если в первой БД изменить варианты ответа на какой-либо закрытый вопрос (например, добавить новый вариант ответа), то приведенный выше запрос не даст информации о том, почему в ряде анкет совершенно отсутствуют такие варианты ответа. Если же

обратиться ко второй БД, то добавленные временные поля как раз и позволяют решить эту задачу, указывая, в какой период существовал тот или иной вариант ответа, хотя сам запрос не изменяется.

Усложнение запросов происходит, когда надо извлечь, например, вопрос и все варианты ответа на него. Запрос к первой БД будет иметь вид:

```

SELECT * FROM
    closed_answers, close_answers_options
WHERE
    closed_answers.close_answers_options_id = close_answers_options.close_answers_options_id
    
```

Данный запрос вернет все варианты независимо от того, актуальны они сейчас или нет. Этот же запрос справедлив и при обращении ко второй БД, однако применение дополнительных вре-

менных полей за счет некоторого усложнения запроса позволяет уточнить такие требования, как выбор всех записей, актуальных в настоящий момент. Усложненный запрос будет иметь вид:

```

SELECT * FROM
    closed_answers, close_answers_options
WHERE
    closed_answers.close_answers_options_id = close_answers_options.close_answers_options_id
AND
    close_answers_options.close_answers_options_valid_end > NOW() AND
    close_answers_options.close_answers_options_valid_begin < NOW()

```

В этом запросе добавляются два дополнительных условия, ограничивающие период актуальности вариантов ответа.

Таким образом, можно сделать ряд выводов. Во-первых, определение периодов актуальности для всех элементов анкеты устраняет неоднозначности, касающиеся формулировок вопросов, мест проведения анкетирований и т. п. Во-вторых, добавление дополнительной информации в БД расширяет возможности анализа, позволяя извлекать и анализировать данные в динамике. В-третьих, темпоральную БД значительно проще связать с событиями, непосредственно не отраженными в

ней, но привязанными к определенным временным интервалам.

В целом темпоральная база данных представляется как востребованное в настоящее время расширение реляционных БД.

#### ЛИТЕРАТУРА

1. Образование, наука, бизнес: особенности регионального развития и интеграции: сб. трудов Всероссийской науч.-практ. конф. – Череповец: ИМИТ СПбГУ, 2005. – 344 с.
2. Порай, Д.С. Реализация концепции темпоральной базы данных средствами реляционной СУБД [Электронный ресурс] / Д.С. Порай, А.В. Соловьев, Г.В. Корольков. – Режим доступа: <ftp://ftp.dol.ru/pub/users/cgntv/download/sbornic/sbornic5/Doc8.doc>
3. Snodgrass, R. T. Developing time-oriented database application in SQL / R.T. Snodgrass. – San-Francisco: Morgan Kaufmann Publishers, 2000. – 528 p.

УДК 331.5(470.12)

**Анна Сергеевна Кельсина**

## ОСОБЕННОСТИ ТРУДОВОГО ПОВЕДЕНИЯ СОВРЕМЕННЫХ ПОДРОСТКОВ

В статье представлены результаты опроса школьников 10 – 11 классов, обучающихся в Научно-образовательном центре ВНКЦ ЦЭМИ РАН, проведенного с целью изучения особенностей поведения и ценностных предпочтений подростков в сфере труда и занятости. Обозначены основные проблемы и трудности, с которыми они сталкиваются на рынке труда.

Формирование и развитие трудового потенциала молодежи напрямую

зависит не только от образования, уровня и качества жизни, но и от мотивационных установок, трудовых ценностей.

Изучение их, как и жизненных планов молодежи, позволяет понять направленность и содержание ее трудовой активности, способствует эффективности разрабатываемых мероприятий в сфере образования, труда и занятости.



А.С. КЕЛЬСИНА –  
инженер-исследователь  
ВНКЦ ЦЭМИ РАН