



**МИН ТХЕТ ТИН**

**МЕТОДИКА ФОРМИРОВАНИЯ РЕЛЯЦИОННЫХ ТАБЛИЦ  
НА ОСНОВЕ ИНФОРМАЦИИ ТАБЛИЧНОГО ВИДА**

Специальность 05.13.11 – Математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования (ФГБОУ ВПО) «Московском государственном техническом университете им. Н.Э. Баумана» на кафедре «Компьютерные системы и сети» ИУ-6.

Научный руководитель:	Доктор технических наук, доцент, профессор кафедры «Информатика и системы управления», МГТУ имени Н.Э. Баумана Брешенков Александр Владимирович
Официальные оппоненты:	Доктор технических наук, профессор, профессор кафедры «Информационные технологии в управлении» ГАОУ ВПО Московский городской университет управления Правительства Москвы (МГУУ) Данчул Александр Николаевич. Кандидат технических наук, доцент, начальник отдела ЗАО «Всесоюзный институт волоконно-оптических систем связи и обработки информации» Самарев Роман Станиславович.
Ведущая организация:	Открытое акционерное общество «Государственный научно-исследовательский институт приборостроения» (ОАО «ГосНИИП»)

Защита диссертации состоится « 27 » октября 2014г. в 10.00 часов на заседании диссертационного совета Д212.125.01 при Московском авиационном институте (национальном исследовательском университете) – МАИ по адресу: 125993, г. Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4.

С диссертацией можно ознакомиться в библиотеке Московского авиационного института (национального исследовательского университета) – МАИ

Отзывы, заверенные печатью, просьба высылать по адресу: 125993, г. Москва, А-80, ГСП-3, Волоколамское шоссе, д.4, МАИ, Ученый совет МАИ

Автореферат разослан « \_\_\_\_ » \_\_\_\_\_ 2014 г.

Ученый секретарь

диссертационного совета Д212.125.01

кандидат технических наук, доцент



А.В.Корнеевкова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность проблемы.** В настоящее время трудно переоценить значение компьютерных информационных систем. А коль скоро базы данных (БД) являются ядром информационных систем, в полной мере это относится и к БД. Это детально и убедительно доказывается в соответствующей научно-популярной и технической литературе. Более того, в паспорте специальности 05.13.11 (Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей) отмечается:

- необходимость разработки и исследования в области программных средств организации и управления обработкой данных и знаний;

- необходимость создания прикладного математического обеспечения, программных средств автоматизации разработки программ;

- актуальность разработки программных средств обработки данных и знаний в ВМ, ВК и КС;

- актуальность разработки методов проектирования систем управления базами данных (СУБД) и базами знаний (СУБЗ), в том числе распределенными СУБД и СУБЗ.

Собственно понятие информации глобальное и охватывает все сферы человеческой деятельности от вербального общения между людьми до работы в интернете. А данные – это информация, представленная в регламентированном виде. К сожалению, не всю информацию можно строго регламентировать и использовать в реляционных базах данных (РБД). Поэтому работы в этом направлении представляют интерес. В диссертации рассматривается информация табличного вида (ИТВ). В качестве примеров ИТВ можно назвать электронные таблицы, таблицы текстовых процессоров, HTML-таблицы и др. Практически на всех предприятиях накоплены значительные объемы ИТВ и они заинтересованы в использовании преимуществ РБД. Представления такого рода информации близки к представлению данных в РБД, и поэтому в принципе процесс преобразования ИТВ в формат РБД можно формализовать и исключить возможные ошибки при проектировании РБД с нуля.

Достаточно большой объем работы в области проектирования РБД на основе использования ИТВ проделал Брешенков А.В. Однако, несмотря на глубокую теоретическую и практическую проработку проблемы, в его работах не рассматриваются некоторые задачи преобразования. В частности:

- рассмотрены не все возможные виды подзаголовков в ИТВ;
- не рассмотрены гибридные подзаголовки;
- в качестве атрибутов, которые входят в первичный ключ, анализировалось не более 2-х;

- связи между таблицами рассмотрены для ключевых полей, включающих только один атрибут;

– не проанализировано одно из требований минимальности первичного ключа – никакая часть первичного ключа не должна быть уникальной;

– не проведены детальные исследования по поводу выявления внешних ключей в ИТВ.

К настоящему времени выполнен значительный объем научных исследований, посвященных проектированию реляционных баз данных (РБД). Среди них можно назвать работы Е. Ф. Кодда, К. Дж. Дейта, Гэри Хансена, Джэймса Хансена, Ульмана Дж., Чена Р. Р., Райана Стивенса, Рональда Плю, Дэйва Энсора, Тихомирова Ю.В., Григорьева Ю.А., Ревункова Г.И. и других. Но в этих работах, как правило, методы проектирования РБД основываются на анализе предварительно разработанных схем отношений, когда данных, как таковых, еще нет.

**Проблема** заключается в отсутствии комплекса способов, алгоритмов, средств и методики, ориентированных на преобразование заполненных ИТВ в реляционные таблицы (РТ).

**Предметом исследования** являются модели, способы и методика проектирования РБД на основе использования существующей, заполненной ИТВ, а также компоненты математического, лингвистического, информационного и программного обеспечений методики.

**Цель и основные задачи исследования.** Целью работы является разработка в рамках предложенной автором методики теоретических и практических основ формирования РТ на базе ИТВ, улучшение качественных и количественных характеристик существующих средств для:

- обработки всех возможных видов подзаголовков в ИТВ;
- исключения гибридных подзаголовков;
- исключения повторяющихся заголовков и подзаголовков;
- приведения значений каждого атрибута ИТВ к одному типу;
- использования в качестве первичных ключей произвольного числа атрибутов;
- анализа требования минимальности первичного ключа;
- выявления внешних ключей в ИТВ.

**Методы исследования.** При разработке формальных моделей, способов и методики в диссертации использованы теория отношений, теория множеств, реляционная алгебра, алгебра логики, теория алгоритмов, аппарат сетей Петри.

**Научную новизну** работы определяет концепция и теоретические основы формирования РТ на базе ИТВ, которые воплощены в соответствующую методику.

**Научные результаты, выносимые на защиту:**

- Исследована проблема проектирования РБД на основе ИТВ.
- Уточнена модель реляционных таблиц.
- Построена модель информации табличного вида.
- Предложен новый способ преобразования ИТВ в РТ.
- Предложен способ назначения ключевых полей в заполненных таблицах ИТВ.

- Разработана методика формирования РТ на основе ИТВ.

### **Достоверность научных положений, рекомендаций и выводов**

Обоснованность научных положений, рекомендаций и выводов, изложенных в работе, определена корректным использованием современного математического аппарата. Достоверность положений и выводов диссертации подтверждена положительными результатами внедрения в учебный процесс МГТУ им. Н.Э. Баумана.

### **Практическая ценность и реализация результатов работы**

Научные результаты, полученные в диссертации, доведены до практического использования. Методика, способы, а также программные средства могут быть использованы при решении задач проектирования РБД на основе использования ИТВ.

Содержание отдельных разделов и диссертации в целом было изложено и получило одобрение:

- на Российских НТК и семинарах (2011 - 2014 г.г.);
- на заседании кафедры “Компьютерные системы и сети” МГТУ им. Н.Э. Баумана.

Совокупность научных положений, идей и практических результатов исследований составляет оригинальное направление в области проектирования реляционных баз данных.

По результатам выполненных исследований опубликовано 11 научных работ.

Диссертационная работа состоит из введения, четырех глав и заключения, опубликованных на 149 страницах машинописного текста, содержит 77 рисунков, 23 таблицы, список литературы из 100 наименований и 2 приложений.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** показана актуальность решаемой проблемы, сформулированы цель и задачи исследования, приведено краткое описание содержания глав диссертации.

**В первой главе** «Анализ проблем разработки методики формирования реляционных таблиц на основе использования информации табличного вида» описан подход к проблеме формирования формирования реляционных таблиц в рамках современной методологии проектирования РБД, сформулированы его достоинства и недостатки. Определено понятие ИТВ и сформулированы мотивы разработки методики преобразования ИТВ в реляционное представление. Сформулированы проблемы разработки методики формирования реляционных таблиц на основе использования ИТВ.

Проектирование РБД в соответствии с традиционной методологией, включает в себя 4 этапа: формулировка и анализ требований, инфологическое проектирование, даталогическое проектирование, физическое проектирование.

По определению реляционная модель данных (РМД) некоторой предметной области представляет набор отношений, изменяющихся во времени. Для однозначного понимания терминологии и сути данной работы оправдано привести таблицу элементов реляционной модели данных (табл. 1).

Таблица 1

Элемент реляционной модели	Форма представления
Отношение	Таблица
Схема отношения	Строка заголовков столбцов таблицы
Кортеж	Строка таблицы
Сущность	Описание свойств объекта
Атрибут	Заголовок столбца таблицы
Домен	Множество допустимых значений атрибута
Значение атрибута	Значения поля в записи
Первичный ключ	Один из нескольких атрибутов
Тип данных	Тип значений элементов таблицы

Основным понятием РМД является отношение, которое представляет собой подмножество декартового произведения доменов  $D_1, D_2, \dots, D_k$  вида:

$$D = D_1 \times D_2 \times \dots \times D_k, \text{ где}$$

$$D_1 = (d_{11}, d_{12}, \dots, d_{1i}, \dots, d_{1m1}); D_2 = (d_{21}, d_{22}, \dots, d_{2i}, \dots, d_{2m2}); \dots; D_k = (d_{k1}, d_{k2}, \dots, d_{ki}, \dots, d_{kmn}).$$

Домен – множество элементов, типы которых могут не совпадать.

Отношение  $R$  представляется следующим образом:

$$R \subseteq D = D_1 \times D_2 \times \dots \times D_k \quad (1).$$

Рассмотрены понятия ключевых полей и обеспечения целостности данных. В работах, посвященных теории проектирования РБД, дается определение первичных ключей, обосновывается их необходимость, формулируются требования к ним и определяются свойства внешних ключей. Обосновано утверждение о том, что при наличии только схемы данных непросто безошибочно выбрать атрибуты, удовлетворяющие этим требованиям.

Ряд специалистов в области реляционных баз данных не включают в качестве требований к РТ наличие первичных ключей. Это спорный подход.

Целостность данных в основном рассматривается в двух аспектах – целостность сущностей и целостность согласования. Требование целостности сущностей можно сформулировать так: сущности реального мира должны быть различимы. То есть ни одна запись в таблице не должна повторяться. В противном случае возникнет множество проблем в процессе обработки данных. Целостность согласования: ссылаться можно только на те данные, которые существуют. Нарушение этого правила может привести к трудно разрешимым проблемам в процессе функционирования БД. Обеспечение целостности данных связано с правильным назначением и использованием первичных и внешних ключей.

Рассмотрены понятия **нормализация и семантическое моделирование**. **Нормализация отношений** – аппарат ограничений на формирование отношений, который позволяет устранить избыточность БД, обеспечивает непротиворечивость хранимых данных.

**Семантическое моделирование**. Разработчики БД обычно обладают весьма ограниченными сведениями о смысле хранящихся в них данных. Поэтому широкое

распространение получил метод семантического моделирования “сущность-связь” или ER-модель. Связи в модели “сущность-связь” могут иметь тип “один к одному”, “один ко многим”, “многие к одному” и “многие ко многим”.

**Определено понятие ИТВ.** Представление информации в табличном виде настолько удобно, что это определило появление целого класса систем, ориентированных на работу с табличной информацией – электронных таблиц. Форма представления ИТВ может быть самой различной: на бумаге, в виде файла текстового редактора, в виде электронных таблиц и др.

Всесторонний анализ ИТВ показал, что в общем случае она обладает следующим свойствами.

1. ИТВ – это информация, которая воспринимается ее потребителями, как таблицы.
2. В ИТВ могут отсутствовать разделители строк и разделители столбцов.
3. Элементы данных таблицы могут размещаться в нескольких строках.
4. Типы элементов данных одного столбца могут не совпадать.
5. Заголовки ИТВ могут включать в себя подзаголовки нескольких уровней.
6. В ИТВ могут быть задействованы внутренние подзаголовки.
7. В ИТВ имена заголовков могут совпадать.
8. В ИТВ возможны пустые строки.
9. В ИТВ могут отсутствовать заголовки столбцов.
10. В ИТВ возможно дублирование строк.

В качестве примера на рис.1 приведен фрагмент реальной ИТВ.

		Продажа модернизации - месяц					Продажа модернизации - YTD					Коммерческие предложения - месяц					Кв			
		Объем		Единицы		Цели		Объем		Единицы		Цели		План		Факт		Цели		План
Region/branch		План	Факт	План	Факт	Объем 000 руб.	ед.	План	Факт	План	Факт	Объем 000 руб.	ед.	ед.	Объем 000 руб.	ед.	Объем 000 руб.	ед.	Объем 000 руб.	ед.
<b>MSR</b>		<b>12500,0</b>	<b>52,6</b>	<b>66,7</b>	<b>1,0</b>	<b>14750,0</b>	<b>77,2</b>	<b>1375,0</b>	<b>52,6</b>	<b>6,0</b>	<b>1,0</b>	<b>3150,0</b>	<b>15,0</b>	<b>75,8</b>	<b>16666,7</b>	<b>35,0</b>	<b>8379,0</b>	<b>178,0</b>	<b>39166,7</b>	<b>62,5</b>
<b>RU-1</b>		250,0	0,0	1,0	0,0	600,0	2,8	750,0	0,0	3,0	0,0	1800,0	8,5	11,4	2500,0	26,0	7309,0	27,3	6000,0	34,1
Б.Курган		0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Д.Самоевское		83,3		0,3		158,3	0,8	250,0	0,0	1,0	0,0	475,0	2,3	3,8	833,3			7,2	1583,3	11,4
А.Трофимов		83,3		0,3		150,0	0,8	250,0	0,0	1,0	0,0	450,0	2,3	3,8	833,3	7,0	1973,0	6,8	1500,0	11,4
О.Валеев (лев)		41,7		0,2		145,8	0,7	125,0	0,0	0,5	0,0	437,5	2,0	1,9	416,7	11,0	3554,0	8,6	1458,3	5,7
А.Моисеев (лев)		41,7		0,2		145,8	0,7	125,0	0,0	0,5	0,0	437,5	2,0	1,9	416,7	8,0	1782,0	8,6	1458,3	5,7
<b>RU-2</b>		208,3	52,6	1,0	1,0	450,0	2,2	625,0	52,6	3,0	1,0	1350,0	6,5	9,5	2083,3	9,0	1070,0	20,5	4500,0	28,4
А.Михайлов		0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
С.Савицкий		104,2		0,5		154,2	0,8	312,5	0,0	1,5	0,0	462,5	2,3	4,7	1041,7			7,0	1541,7	14,2
В.Левин		104,2	52,6	0,5	1,0	154,2	0,8	312,5	52,6	1,5	1,0	462,5	2,3	4,7	1041,7	9,0	1070,0	7,0	1541,7	14,2
Валеев-1		0,0		0,0		141,7	0,7	0,0	0,0	0,0	0,0	425,0	2,0	0,0	0,0			8,4	1416,7	0,0
<b>RU-3</b>		250,0	0,0	1,0	0,0	600,0	2,8	0,0	0,0	0,0	0,0	0,0	0,0	11,4	2500,0	0,0	0,0	27,3	6000,0	0,0
А.Былкин		0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Рис.1. Фрагмент реальной ИТВ.

Разработан способ преобразования ИТВ в РТ. Анализ проблемы показал, что для этого необходимо разработать следующие взаимосвязанные алгоритмы:

- исключение заголовков внутри таблиц;
- исключение подзаголовков внутри таблицы;
- исключение сложных подзаголовков;
- исключение пустых подзаголовков;

- избавление от вертикальных подзаголовков;
- приведение значений атрибутов к одному типу;
- исключение дублирования записей;
- исключение пустых записей.

Построена схема иерархии разрабатываемых алгоритмов (рис. 2).



Рис. 2. Схема иерархии разрабатываемых алгоритмов

Во второй главе «Способ преобразования заполненных нереляционных таблиц в реляционные таблицы» разработаны модели объектов исследования (РТ и ИТВ), способ преобразования таблиц ИТВ к реляционному виду. В рамках способа рассмотрены алгоритмы приведения значений каждого атрибута заполненной таблицы к одному типу, исключения дублирования записей в ИТВ, избавления от сложных атрибутов и исключения простых и сложных подзаголовков.

**Основные черты формальной модели РТ следующие.** Реляционная таблица (RT) представляется множеством

$RT = \{Z, D\}$ , где  $Z$  – множество заголовков,  $D$  – множество данных.

$z = \{z_1, \dots, z_i, \dots, z_n\}$ ,  $i = 1, n$ ;  $n \geq 1$ , где  $n$  – степень множества заголовков.

Должно быть обеспечено условие  $z_i \neq z_m$ ,  $i = 1, n$ ;  $m = 1, n$ ;  $i \neq m$ ,

где  $n$  – степень множества заголовков, т.е. недопустимо совпадение заголовков.

$D = \{SD\}$ , где  $SD$  – множество строк данных.

$SD = \{SD_1, \dots, SD_i, \dots, SD_n\}$ ,  $i = 1, n$ ;  $n \gg 1$ ,

где  $n$  – мощность множества строк данных.

$SD_i = \{ED_{i1}, \dots, ED_{ij}, \dots, ED_{ik}\}$ ,  $j = 1, k$ ;  $k \geq 1$ ,

где  $k$  – степень множества  $i$ -ой строки данных,  $ED_{ij}$  – элемент данных.

**В РТ не должно быть пустых заголовков:  $Z_i \neq \emptyset$ .**

**В РТ не должно быть пустых строк:  $SD_i \neq \emptyset$ .**



***В РТ содержимое столбца не может использоваться как подзаголовки:  $SD_i \neq zn$ .***

***В РТ не должно быть внутренних подзаголовков:  $SD_i \neq z$ .***

***В РТ не должно быть сочетаний различного типа подзаголовков:  $\neg((Sd_i = zn) \& (SD_i = z))$ .***

Жирным курсивом выделены уточнения модели РТ, сделанные автором.

***Модель ИТВ*** в известной литературе не рассматривается. Рассмотрим эту модель. ИТВ представляются множеством  $DT = \{Z, D\}$ , где  $Z$  – множество заголовков ИТВ,  $D$  – множество данных, соответствующих заголовкам.

$Z = \{Z_1, \dots, Z_i, \dots, Z_n\}$ ,  $i = 1, n$ ;  $n \geq 1$ , где  $n$  – где  $n$  число заголовков.

Допустима ситуация, когда  $Z_i = Z_m$ ,  $i = 1, n$ ;  $m = 1, n$ ;  $i \neq m$ , где  $n$  – степень множества заголовков, т.е. возможно полное совпадение заголовков.

$D = \{SD, Z\}$ , где  $SD$  – множество строк данных.

Такого рода представление  $D$  допускает наличие нескольких заголовков и подзаголовков 1-го и 2-го уровней, расположенных в области данных.

$SD = \{SD_1, \dots, SD_i, \dots, SD_n\}$ ,  $i = 1, n$ ;  $n \gg 1$ , где  $n$  – мощность множества строк данных.

$SD_i = \{ED_{i1}, \dots, ED_{ij}, \dots, ED_{ik}\}$ ,  $j = 1, k$ ;  $k \geq 1$ , где  $k$  – степень множества элементов данных  $i$ -ой строки данных;  $ED_{ij}$  – элемент данных. В ИТВ могут быть пустые заголовки:  $Z_i = \emptyset$ . В ИТВ могут быть пустые строки:  $SD_i = \emptyset$ .

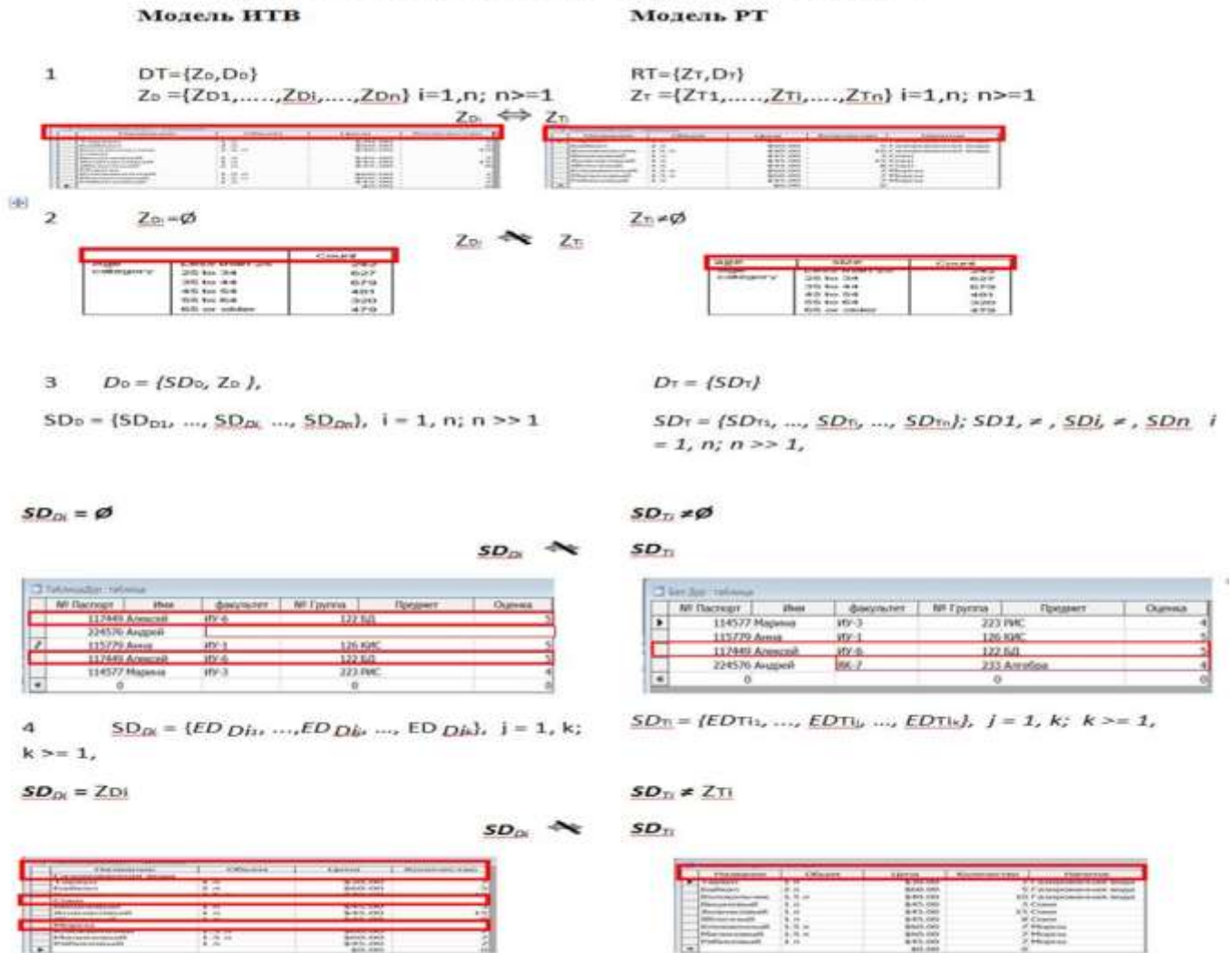
В РТ основные типы полей: числовой, текстовый, дата-время. В нереляционных таблицах в одном и том же столбце могут храниться данные различных типов. Это недопустимо в РТ. Поэтому необходимы средства преобразования каждого элемента каждого столбца таблиц ИТВ к одному типу.

В РТ недопустимо дублирование записей, а в ИТВ это возможно. В связи с этим необходимы средства исключения дублирования записей в ИТВ.

На рис. 3 представлены результаты сравнительного анализа модели информации табличного вида и модели реляционных таблиц. Этот анализ позволил формально описать сходства и различия моделей, и как следствие, сходство и различие объектов, которые соответствуют этим моделям.

Так как основной задачей работы является разработка алгоритмов, способов, методики преобразования ИТВ в РТ, то каждая позиция сравнительного анализа позволяет определить какие задачи необходимо решать в ходе разработки методики преобразования ИТВ в РТ. Упрощенно разработанная методика представляет собой процесс преобразования объектов, соответствующих модели ИТВ к объекту, соответствующему модели РТ. А совокупность связанных между собой реляционных таблиц и является реляционной базой данных. Следует обратить внимание на то, что методика проектирования систем управления базами данных в полном объеме в работе не рассматривается, а лишь представлена основными характерными запросами, которые формируются в ходе преобразования.

# Сравнительный анализ моделей ИТВ и РТ



$Z_D$  – множество заголовков ИТВ,  $Z_T$  – множество заголовков РТ,

$SD_D$  – множество строк данных ИТВ,  $SD_T$  – множество строк данных РТ

Рис.3. Результаты сравнительного анализа модели ИТВ и модели РТ.

**Исключение сложных атрибутов и подзаголовков.** В ИТВ могут встречаться подзаголовки трех типов: внешние подзаголовки, внутренние подзаголовки и подзаголовки-столбцы, а также могут быть различные сочетания подзаголовков. В РТ подзаголовки недопустимы. На основе сравнительного анализа предложенных моделей РТ и ИТВ в работе разработаны алгоритмы исключения подзаголовков всех типов. Из соображений ограниченного объема автореферата приводится самый простой алгоритм – алгоритм исключения внутренних подзаголовков.

```

COUNTER = 0
FOR s = 1 to m
    COUNTER1 = 0
    FOR f = 1 to k
    
```

```

    IF ask= NULL THEN COUNTER1 = COUNTER1 + 1
NEXT f
    IF COUNTER1 = k-1 THEN COUNTER= COUNTER+1
NEXT s
IF COUNTER < 2 THEN EXIT
Формирование двух отношений
R0 = R(A1, ..., Ai, ..., Ak)+R(KR)
COUNTER = 0
FOR s = 1 to m
    COUNTER1 = 0
    FOR f=1 to k
        IF ask= NULL THEN COUNTER1 = COUNTER1 + 1
    NEXT f
    IF COUNTER1 = k-1 THEN
        COUNTER = COUNTER + 1
        C(R2COUNTER,1)=COUNTER
        C(R2COUNTER,2)= ask
        DELETE * FROM R0 WHERE (A1= ask)
    ELSE
        C(R0s,1)=COUNTER
    END IF
NEXT s

```

Здесь  $m$  – мощность отношения  $R$ ,  $k$  – степень отношения  $R$ .

Выражение  $R0 = R(A1, \dots, Ai, \dots, Ak)+R(KR)$  означает добавление к  $R$  атрибута с именем  $KR$ .

Выражение  $C(R2COUNTER,1)$  означает значение элемента  $R2$  в строке  $COUNTER$  и 1-ом столбце.

$ask$  – счетчик пустых полей.

Выражение  $C(R0s,1)$  означает значение элемента  $R0$  в строке  $s$  и 1-ом столбце.

По сути алгоритм организует поиск записей, в которых все значения кроме 1-го – пустые (Null). Найденные подзаголовки удаляются, но предварительно преобразуются в новый атрибут таблицы.

**Проблема приведения значений атрибутов заполненных таблиц к одному типу** решена двумя способами – с использованием существующих инструментальных средств, в частности Access, для таблиц малой размерности (менее 200 записей) и с использованием разработанных алгоритмов и соответствующих процедур для таблиц большой размерности.

**Исключение дублирования записей** в основном, как показано в работе, хорошо обеспечивается существующими инструментальными средствами, которые органично

интегрируются в разработанное приложение. В качестве основного критерия использовано время выполнения запроса. Оно исчисляется долями секунды.

При этом задействуются 3-и запроса типа:

```
SELECT DISTINCT Сотрудники.Фамилия, Сотрудники.Имя, Сотрудники.Отчество  
INTO Временная FROM Сотрудники;
```

```
DELETE Сотрудники.* FROM Сотрудники;
```

```
INSERT INTO Сотрудники SELECT Временная.* FROM Временная;
```

**Суть способа преобразования ИТВ в РТ состоит в преобразовании объектов, соответствующих модели ИТВ, к объектам, соответствующим модели РТ, на основе использования предложенного комплекта алгоритмов и разработанной методики.**

В третьей главе «Способ назначения ключевых полей в заполненных таблицах» разработан способ назначения ключевых полей в заполненных реляционных таблицах, который по сравнению с традиционными способами позволяет снизить трудоемкость и достигнуть наилучшего решения при выполнении данной проектной процедуры.

В работах Брешенкова А.В. предложен способ назначения ключевых полей в ИТВ, который вполне приемлем для использования. Однако в нем учтены не все особенности ключевых полей. В частности:

- рассматривается возможность включения в первичный ключ только 2-х атрибутов;
- не полностью учитывается требование минимальности первичного ключа;
- не до конца прояснены вопросы формирования первичных ключей из нескольких атрибутов;
- мало освещены вопросы назначения внешних ключей;
- назначение первичных ключей не рассматривается как неотъемлемая задача преобразования ИТВ в РТ.

Ниже сформулированы требования к ключевым полям.

**Уникальность.** Пусть имеется отношение R:

$R=(A_1, \dots, A_i, \dots, A_m, \dots, A_k)$ ,  $i = \overline{1, k}$ , где  $k$  – степень отношения;  $A_i$  – атрибут отношения.

$A_i = \{e_{i_1}, \dots, e_{i_j}, \dots, e_{i_n}\}$   $j = \overline{1, n}$ , где  $n$  – мощность отношения,  $e_{i_j}$  –  $j$ -й элемент атрибута  $A_i$ ;

$A_m = \{e_{m_1}, \dots, e_{m_j}, \dots, e_{m_n}\}$   $j = \overline{1, n}$ , где  $n$  – мощность отношения,  $e_{m_j}$  –  $j$ -й элемент атрибута  $A_m$ .

Необходимо найти такие атрибуты  $A_i, \dots, A_m$  чтобы обеспечилась истинность выражения:  
 $\text{concat}(e_{i_1}, \dots, e_{m_1}) \neq, \dots, \neq \text{concat}(e_{i_j}, \dots, e_{m_j}) \neq, \dots, \neq \text{concat}(e_{i_n}, \dots, e_{m_n})$  (1)

Необходимо найти такое сочетание атрибутов, чтобы конкатенация их значений была уникальна. При этом:

- проверяемый кортеж атрибутов может включать несколько атрибутов;
- число возможных сочетаний атрибутов может быть очень большим – это зависит от степени отношения (общего числа атрибутов в отношении);
- ключевой атрибут может быть только один;

– может не найтись таких атрибутов, которые обеспечивают истинность выражения (1), в этом случае назначают суррогатный ключ.

**Минимальность.** Минимальность ключевого поля рассматривается в двух аспектах.

**В первой части требования** во главу угла становится объем памяти, который необходим для хранения значений атрибутов, входящих в первичный ключ. Поэтому самая очевидная целевая функция – минимальное число атрибутов, входящих в первичный ключ:  $\min|A_1, \dots, A_i, \dots, A_k|$ ,  $i = \overline{1, k}$ , где  $k$  – число атрибутов, входящих в первичный ключ;  $A_i$  – атрибут отношения, входящий в первичный ключ.

Строго говоря, целевая функция следующая:  $\min(\text{Length}(A_i) + \dots + \text{Length}(A_k))$ .

**Во второй части требования** под минимальностью первичного ключа подразумевается отсутствие в составе ключа атрибута, значения которого уникальны. Пусть первичный ключ  $K$  представлен множеством атрибутов:  $K = (A_1, \dots, A_i, \dots, A_j, \dots, A_k)$ ,  $i = \overline{1, k}$ , где  $k$  – число атрибутов, входящих в первичный ключ;  $A_i$  –  $i$ -й атрибут отношения, входящий в первичный ключ. Тогда должно быть истинно выражение  $\neg (e_{i_1}, \dots, \neq, \dots, e_{i_j}, \dots, \neq, \dots, e_{i_n})$ .

**Алгоритмы назначения первичных ключей в заполненных таблицах.** Предложены неформальные (описательные) и формальные алгоритмы назначения первичных ключей. Из соображений объема приводится только один неформальный алгоритм – алгоритм поиска ключа, включающего пару атрибутов:

$MKA^2 = \emptyset$ , где  $MKA^2$  – множество пар атрибутов, которые претендуют на роль первичного ключа.

Пусть имеется отношение  $R$ :  $R = (A_1, \dots, A_i, \dots, A_k)$ ,  $i = \overline{1, k}$ , где  $k$  – степень отношения;  $A_i$  – атрибут отношения.  $MPA = \emptyset$ .

Ищутся все возможные сочетания пар атрибутов и запоминаются в массив  $MPA$ :

```
Count=0
For i = 1 to k-1
  For j = i+1 to k
    Count = Count + 1
    S = Concat (Ai , Aj)
    MPA (Count) = MPA + S
  Next j
Next i
```

Таким образом, в массиве  $MPA$  сформируются все возможные пары атрибутов, а в счетчике  $Count$  хранится их количество.

Проверяются все пары на уникальность.

$MUP = \emptyset$  /\* Массив пар атрибутов, представляющих собой атрибуты, все соответствующие пары значений которых уникальны \*/

```
Count1 = 0
```

```

For i = 1 to Count
  S = MPA(Count)
  /* По сути S представляет собой пару атрибутов (Ai, Aj)
  Ai = (ei1 ... eim), где ei1 – 1-й элемент домена с атрибутом Ai, eim – m-й элемент домена с
  атрибутом Ai.
  Aj = (ej1 ... ejm), где ej1 – 1-й элемент домена с атрибутом Aj, ejm – m-й элемент
  домена с атрибутом Aj, m – мощность отношения. */
  For n = 1 to m
    /* Для каждой пары атрибутов (Ai, Aj) выполняется проверка условия
    Concat(ei1, ej1) ≠ ... ≠ Concat(eim, ejm) */
    Next n
    /* Если текущая пара атрибутов имеет все соответствующие пары значений
    уникальными, то эта пара добавляется к массиву пар с уникальными значениями: */

    Count1 = Count1 + 1
    MUP(Count1) = S
  Next i

```

Если претенденты на ключевой атрибут найдены, т.е. ,  $MUP \neq \emptyset$ , то для проверки второго требования минимальности выполняется переход к алгоритму поиска первичного ключа на основе 3-х атрибутов и далее по аналогии. Разработчик на любом шаге поиска может принять решение о назначении суррогатного ключа.

**Алгоритмы назначения внешних ключей в заполненных таблицах.** В диссертации предложены неформальные и формальные алгоритмы назначения внешних ключей в заполненных таблицах. Ниже приведен формальный алгоритм.

П1. Осуществляется поиск всех возможных сочетаний пар таблиц анализируемой БД.

```

/* T=(T1, ..., Ti, ..., Tm, ..., Tk), i =  $\overline{1, k}$ , где k – число прикладных таблиц БД */
MPT =  $\emptyset$  __ Массив всех пар таблиц БД
/* Ищутся все возможные сочетания пар таблиц */
Count=0
For i = 1 to k-1 __ k – число таблиц БД
  For j = i+1 to k
    Count = Count + 1 __ счетчик числа пар таблиц
    S = Concat (Ti, Tj)
    MPT (Count) = S
  Next j
Next i

```

П2. Для каждой пары таблиц из массива МРТ выполняется поиск внешнего ключа

```

F = 0 __ флажок наличия связанных ключей
CountI = 0 __ счетчик совпадений
For i = 1 to Count __ – число пар таблиц БД
  For l = 1 to C1 __ – степень 1-й таблицы из i-й пары
    For m = 1 to C2 __ – степень 2-й таблицы из i-й пары
      For j = 1 to M1 __ – мощность 1-й таблицы из i-й пары
        For k = 1 to M2 __ – мощность 2-й таблицы из i-й пары
          If ei,j = em,k Then Coun1 = Coun1 + 1
        Next k
      Next j
    Next m
  Next l
  If Coun1 >= Max(M1, M2) Then
    Begin
      Print (“Для пары таблиц ”, i, “столбцы”, l, “и”, m)
      Print (“ претендуют на первичный и внешние ключи”)
      CountI = 0
      F = 1
    End
  Next m
Next l
Next i
If F = 0 Then Print (“ внешних ключей не обнаружено не для одной из таблиц”)

```

Следует обратить внимание на то, что атрибуты, предложенные алгоритмом в качестве связанных атрибутов, могут не удовлетворять разработчика, поэтому необходимо предоставить ему возможность окончательного решения.

**Суть способа назначения ключевых полей в ИТВ состоит в комплексном использовании разработанных алгоритмов, которые на основе предложенных моделей РТ и ИТВ обеспечивают преобразования ИТВ в РТ в соответствии с требованиями к ключевым полям в рамках разработанной методики.**

**В четвертой главе «Методика формирования реляционных таблиц на основе информации табличного вида»** решена задача формализации методики формирования РТ на основе ИТВ. Разработана операторная модель и сетевая модель методики. Выполнено ее исследование на основе использования сетевой модели.

**Постановка задачи формализации методики.** Целью работы является разработка методики формирования РТ на основе ИТВ. Методика преобразования ИТВ в РТ – это процесс, в котором активное участие принимает разработчик.

**Операторная модель методики преобразования ИТВ в РТ.** На начальном уровне абстрагирования от большинства компонент человеко-машинной системы схема процесса преобразования ИТВ в РТ может быть проиллюстрирована рис.4.

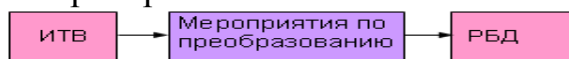


Рис. 4. Исходная схема процесса преобразования ИТВ в РТ.

Введем оператор преобразования ИТВ в РБД – ОП. Результатом применения этого оператора ОП является набор РТ, соответствующих РБД. Причем РБД должна отвечать требованиям непротиворечивости, минимальности, целостности. Схема преобразования модели ИТВ в модель РБД или операторная модель преобразования примет вид рис. 5.

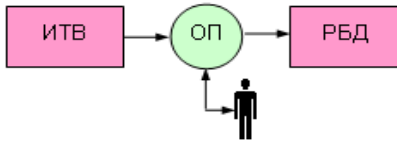


Рис. 5. Исходная операторная модель процесса преобразования ИТВ в РБД

Оператор операторной модели (окружность) соответствует одной или более человеко-машинной процедуре. Состояние операторной модели (прямоугольник) соответствует начальному, промежуточному и окончательному состоянию ИТВ.

В диссертации выполнено 11 итераций построения операторной модели методики преобразования ИТВ в РТ. Модель анализировалась и уточнялась. В результате получена начальная операторная модель, представленная на рис. 6.

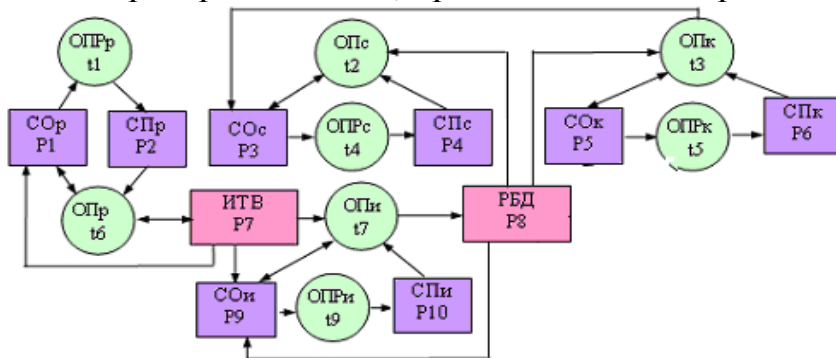


Рис. 6. Начальная операторная модель процесса преобразования ИТВ в РБД

Здесь индексам “и” соответствуют средства импорта, индексам “р” – средства преобразования ИТВ в РТ, индексам “к” – средства преобразования назначения ключевых полей, индексам “с” – средства формирования связей между таблицами, С - системы оценок, ОП - операторы.

В соответствии с моделью после каждого выполнения оператора с помощью автоматизированных средств выполняется проверка промежуточного состояния таблицы ИТВ на соответствие модели РТ. Успешным завершением работы в рамках разработанной модели методики считается полное соответствие таблицы, описываемой моделью ИТВ, таблице, описываемой моделью РТ.

**Мотивация перехода от операторной модели к модели, построенной на основе аппарата сетей Петри.** Операторная модель, построенная на основе анализа отличий моделей ИТВ и РТ, позволила выделить основные компоненты человеко-машинной подсистемы преобразования, определить порядок их использования, сформировать связи между ними. Во избежание ошибок на ранних этапах разработки методики и подсистемы необходимо исследовать следующие ее характеристики:



- выполнение баланса входных и выходных информационных потоков для всех систем принятия решения (устойчивость);
- достижимость всех состояний человеко-машинной системы (живость);
- отсутствие тупиковых ситуаций в процессе преобразования (живость);
- отсутствие ситуаций, когда подсистема приходит в нестационарное состояние и число информационных потоков превышает критическую отметку;
- оценку интегральных временных характеристик подсистемы.

Операторная модель не позволяет выполнить эти исследования.

Предложенное операторное представление методики наиболее близко к системному уровню, а на данном уровне обычно применяют модели систем массового обслуживания или сети Петри. В связи с этим для выявления принципиальных ошибок в методике и ее исследования выбран аппарат сетей Петри. После проведения исследований уточнена операторная модель.

**Формирование сетевой модели.** Сеть Петри состоит из трех элементов: множество мест  $S$ , множество переходов  $T$  и отношение инцидентности  $F$ .

Между набором  $N = (S, T, F)$  сети Петри и набором  $M = (C, O, L)$  операторной модели установлено соответствие таким образом, чтобы между элементами наборов обеспечивалось взаимно однозначное соответствие:  $C \leftrightarrow S$ ;  $O \leftrightarrow T$ ;  $L \leftrightarrow F$ . Как следствие этого соответствия:  $|C| = |S|$ ;  $|O| = |T|$ ;  $|L| = |F|$ .

На рис. 7 приведена сеть Петри, соответствующая начальной операторной модели.

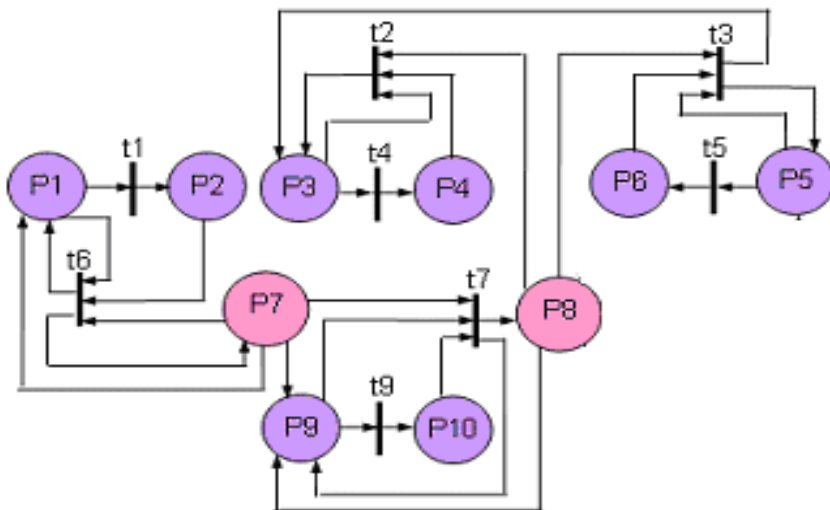


Рис. 7. Сеть Петри, соответствующая начальной операторной модели.

Модели ИТВ, модели РТ, СО, СП поставлены в соответствие положениям сети  $\{P\}$ . Операторам ОП, ОПР поставлены в соответствие переходы сети  $\{t\}$ .

**Исследование устойчивости сети.** Сеть Петри является устойчивой, если она имеет потоковое назначение  $\phi_i > 0$  для каждого  $t_i \in T$ .

Для каждого перехода сети рис. 6 назначен поток  $\varphi_i$ . Для каждого положения  $P_i$  запишем уравнения потоков, которые не должны противоречить друг другу, если данная сеть устойчива. В таблице 2 представлена часть уравнений.

Таблица 2

	t1	t2	t3	t4	t5	t6	t7	t9	
P1	$-\varphi_1$					$-\varphi_6 + \varphi_6$			=0
P2	$+\varphi_1$					$-\varphi_6$			=0
P3		$+\varphi_2 - \varphi_2$	$+\varphi_3$	$-\varphi_4$					=0
P4		$-\varphi_2$		$+\varphi_4$					=0
P5			$+\varphi_3 -$		$-\varphi_5$				=0
P6		$-\varphi_3$		$+\varphi_5$					=0
P7					$-\varphi_6 + \varphi_6$	$-\varphi_7$			=0
P8		$-\varphi_2$	$-\varphi_3$			$+\varphi_7$	$-\varphi_8 + \varphi_8$		=0
P9						$-\varphi_7 + \varphi_7$		$-\varphi_9$	=0
P10						$-\varphi_7$		$+\varphi_9$	=0

После минимизации уравнений оказалось, что часть потоков имеют нулевые значения, а это противоречит требованию устойчивости сети. Поэтому для сетевой модели выполнен ряд итераций преобразования сети и соответствующей ей операторной модели – введены дополнительные состояния и переходы, отражающие реальное положение дел. Достигнута ситуация, когда в уравнениях потоков нет противоречий. Параллельно преобразованиям модифицировалась операторная модель.

**Исследование живости и достижимости сети.** Для этих исследований используется маркированная сеть Петри, которая представляется тройкой  $P = \langle P, T, H \rangle$ , где  $P$  – множество положений;  $T$  – множество переходов;  $H$  – множество маркеров. Динамику функционирования системы можно моделировать перемещением маркеров в сети в соответствии с правилами перехода:  $M'(P) = M(P) - P(t_i) + H(t_i)$ , где  $M(P) = (M(P_1), M(P_2), \dots, M(P_N))$  - разметка сети. На рис. 8 приведен пример перемещения маркеров. Маркеры символизируют состояние системы.

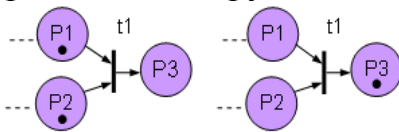


Рис. 8. Пример перемещения маркеров в сети.

В диссертации выполнено перемещение маркеров в сети, при этом проанализировано более 50-и состояний. Это позволило сделать вывод о том, что:

- сеть является живой, т.е. есть принципиальная возможность срабатывания всех переходов;
- сеть является достижимой, т.е. достижимы все положения сети.

В работе выполнена проверка живости и достижимости сети, выявлены ошибки, исправлена сеть Петри и соответствующая операторная модель.

В результате всех преобразований получена сеть Петри свободная от принципиальных ошибок, что исключает принципиальные ошибки в описываемом ею процессе.

После выполненных преобразований сетевой модели и исключения принципиальных ошибок в методике выполнен обратный переход к операторной модели (рис. 9).

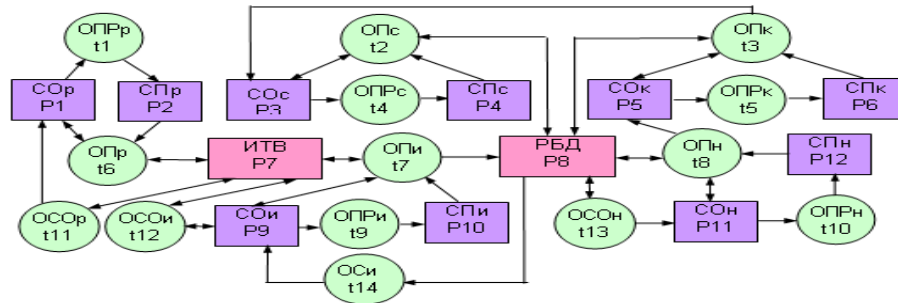


Рис. 9. Окончательная операторная модель методики преобразования ИТВ в РТ.

Разработанная операторная модель методики служит в качестве исходной формализации для процедур подсистемы преобразования ИТВ в РТ, а также как руководство к действиям пользователя подсистемы.

Суть предложенной методики состоит в последовательном использовании разработанных алгоритмов и способов, в которых задействованы предложенные модели РТ и ИТВ, в соответствии с формализованной моделью процесса преобразования.

В приложении 1 «Программная реализация методики формирования реляционных таблиц на основе информации табличного вида» выполнена программная реализация методики формирования реляционных таблиц на основе информации табличного вида. Выполнены экспериментальные исследования временных характеристик процедур подсистемы. На рис. 10. приведено схематичное изображение методики преобразования ИТВ в РТ.

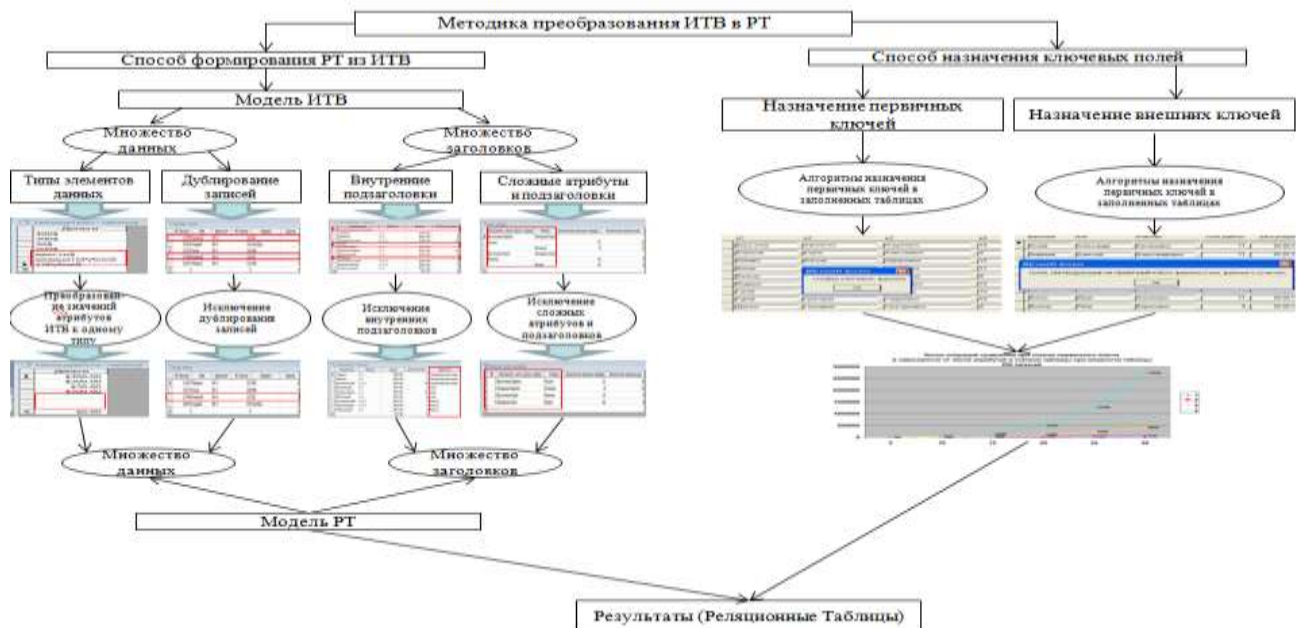


Рис. 10. Схематичное изображение методики преобразования ИТВ в РТ.

**Программная реализация назначения ключевых полей** выполнена для ключевых полей сформированных на базе одного, двух и трех атрибутов таблицы. На рис. 11 приведен фрагмент соответствующего интерфейса



Рис. 11. Фрагмент интерфейса

Ниже приведен фрагмент результата функционирования процедуры, обеспечивающей поиск первичного ключа на базе одного атрибута.

В результате выполнения поиска первичного ключа на базе одного атрибута сформируется сообщение типа рис. 12.

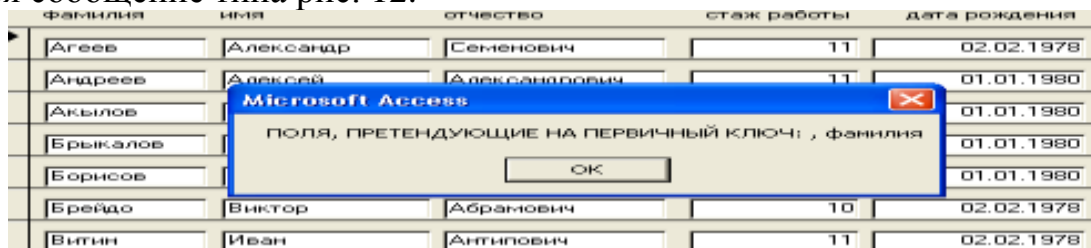


Рис. 12. Сообщение, формируемое после поиска ключа на базе одного атрибута.

В результате выполнения поиска первичного ключа на базе трех атрибутов может сформироваться одно из сообщений типа рис. 13.

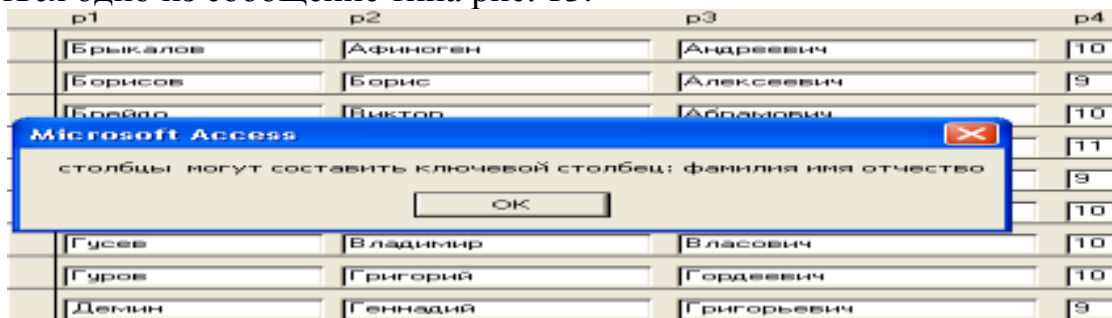


Рис. 13. Одно из сообщений, формируемых в процессе поиска ключа на базе трех атрибутов.

**Оценка вычислительной сложности процедур назначения первичных ключей.**

Для анализируемой таблицы необходимо проверить все сочетания полей. Число возможных сочетаний  $C_n^k = n! / ((n-k)! \cdot k!)$ , где  $n$  – общее число атрибутов таблицы, а  $k$  – количество проверяемых атрибутов на принадлежность к первичному ключу. Выполнены экспериментальные оценки разработанных средств (рис. 14).

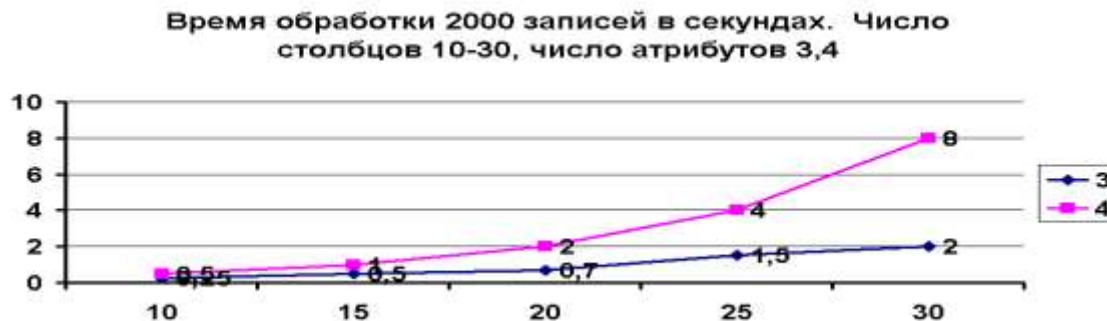


Рис. 14. Графики времени выполнения процедур в зависимости от числа столбцов и числа анализируемых атрибутов.

**Анализ типов атрибутов таблицы.** Все столбцы ИТВ имеют тип строковый. Поэтому задача состоит в том, чтобы проанализировать в каждом столбце все его элементы и по внешнему виду элементов определить его тип. На рис.15 приведен фрагмент интерфейса для анализа типов ИТВ.



Рис 15. Фрагмент интерфейса для анализа типов ИТВ.

После нажатия на соответствующую кнопку для каждого столбца сформируется сообщение о том, сколько полей не являются датами. Пример одного из сообщений приведен на рис. 16.

Антипович	11	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	холост	Б
Борисович	9	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	женат	М
Викторович	10	02			М. Н.Э. Баумана	холост	М
Петрович	11	02			М. Н.Э. Баумана	холост	М
Гордеевич	10	01			М. Н.Э. Баумана	холост	Б
Власович	10	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	женат	М

Рис. 16. Сообщение о том, сколько полей не являются датами.

Следующее сообщение (рис. 17) конкретизирует в каких записях таблицы значения полей не являются датами.

Антипович	11	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	холост	Б
Борисович	9	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	женат	М
Викторович						холост	М
Петрович						холост	М
Гордеевич						холост	Б
Власович	10	02.02.1978	высшее	07.07.1993	МГТУ им. Н.Э. Баумана	женат	М

Рис. 17. Сообщение о том, в каких записях таблицы тип значения – не дата.

В соответствии с полученными сообщениями пользователь может принять решение о том, какие поля необходимо исправлять, а какие нет.

**В заключении** подведены основные итоги выполненных исследований и сформулированы основные выводы и наиболее важные результаты диссертационной работы.

1. Разработаны и формально описаны модели ИТВ и РТ, которые использованы в качестве компонентов методики преобразования ИТВ в РТ.

2. Сформулирована задача и предложен способ преобразования ИТВ в РТ.

3. Сформулирована задача и предложен способ назначения первичных и внешних ключей в заполненных ИТВ.

4. Выполнена формализация методики преобразования ИТВ в РТ, на ее основе проведены исследования и исключены концептуальные ошибки.

5. Разработана автоматизированная подсистема преобразования ИТВ в РТ, проведены экспериментальные исследования.

#### **Основные результаты диссертации отражены в научных работах**

1. Мин Тхет Тин. Электронная цифровая подпись. Современные информационные технологии // Сб. трудов кафедры ИУ-6. – М.: НИИ РЛ МГТУ им. Н.Э. Баумана, 2011. – С. 112–115.

2. Брешенков А.В., Мин Тхет Тин. Исключение внутренних подзаголовков и избавление от сложных атрибутов при преобразовании нереляционных таблиц к реляционному виду // Современные информационные технологии: Сб. трудов кафедры ИУ-6. – М.: НИИ РЛ МГТУ им. Н.Э. Баумана, 2011. – С. 176–183.

3. Брешенков А.В., Мин Т.Т. Вычислительная сложность процедур назначения первичных ключей в заполненных таблицах // Информатика и системы управления в XXI веке: Сб. трудов МГТУ им. Н.Э. Баумана. – М.: МГТУ им. Н.Э. Баумана, 2012.– №9. – С. 136–142.

4. Брешенков А. В., Мин Т. Т. Мотивы разработки метода преобразования информации табличного вида в реляционное представление. // Инженерное образование, 2012. – №3. – 13 с. (Наука и образование: Эл. науч. издание. Номер гос. регистрации 0421200025.)

5. Брешенков А. В., Мин Т. Т. Аналитический обзор традиционного подхода формирования реляционных таблиц с учетом использования существующей информации табличного вида // Инженерное образование, 2012. – №8. – 16 с. (Наука и образование: Эл. науч. издание. Номер гос. регистрации 0421200025.)

6. Брешенков А. В., Мин Т. Т. Модели реляционных таблиц и информации табличного вида. // Инженерное образование, 2012. – №7. – 10 с. (Наука и образование: Эл. науч. издание. Номер гос. регистрации 0421200025.)

7. Брешенков А. В., Мин Т. Т. Алгоритмы назначения первичных ключей в заполненных таблицах. // Инженерное образование, 2012. – №6. – 14 с. (Наука и образование: Эл. науч. издание. Номер гос. регистрации 0421200025.)

8. Брешенков А.В., Мин Т. Т. Преобразование нереляционных таблиц к реляционному виду без использования сложных атрибутов // Вестник Московского государственного технического университета им. Н.Э.Баумана – М., 2012. – №2. – С. 59–60.

9. Мин Тхет Тин, Брешенков А. В., Гудзенко Д. Ю. Назначение внешних ключей в заполненных реляционных таблицах // Современные компьютерные системы и технологии: Сб. трудов каф. ИУ–6 МГТУ им. Н.Э. Баумана, фак. “Компьютерные системы и сети”. М., – 2012.– С. 128–135.

10. Мин Тхет Тин, Брешенков А.В., Гудзенко Д.Ю. Анализ типов атрибутов информации табличного вида // Современные компьютерные системы и технологии: Сб. трудов каф. ИУ-6 МГТУ им. Н.Э. Баумана, фак. “Компьютерные системы и сети”. М., – 2012.– С. 15–23.

11. Мин Тхет Тин, Брешенков А.В., Гудзенко Д.Ю. Назначение внешних ключей в заполненных реляционных таблицах // Современные компьютерные системы и технологии: Сб. трудов каф. ИУ-6 МГТУ им. Н.Э. Баумана, фак. “Компьютерные системы и сети”. М., – 2012.– С. 128–135.