

УДК 378; 681.3.07

Операторная модель процесса преобразования информации табличного вида в реляционные таблицы

Брешенков А. В.^{1,*}, Мин Тхет Тин¹

[*breshenkov@rambler.ru](mailto:breshenkov@rambler.ru)

¹МГТУ им. Н.Э. Баумана, Москва, Россия

В рамках международного научного конгресса "Наука и инженерное образование. SEE-2016", II международная научно-методическая конференция «Управление качеством инженерного образования. Возможности вузов и потребности промышленности» (23-25 июня 2016 г., МГТУ им. Н.Э. Баумана, Москва, Россия).

В статье введено понятие "Информация табличного вида". Перечислены требования к реляционным таблицам (РТ). Обоснована актуальность автоматизированного преобразования информации табличного вида (ИТВ) в РТ и предложена методика преобразования. Выделены задачи преобразования. Средствам решения этих задач поставлены в соответствие операторы. Определены связи между операторами, которые необходимо реализовать для отражения методики преобразования в виде операторной модели. Построена операторная модель преобразования ИТВ различного вида в РТ, на основании которой может быть предложена методика автоматизированного формирования РТ на основе использования заполненных ИТВ. Модель позволяет определить состав процедур и связи между ними для реализации полноценной программной системы.

Ключевые слова: модель, процесс, реляционные таблицы, методика, информация табличного вида

Введение

Информация - понятие, охватывающее все сферы человеческой деятельности от вербального общения между людьми до работы в интернете. В статье рассматривается расширение модели информации табличного вида (ИТВ), предложенной в работах [1, 2]. Представление такого рода информации в табличном виде близко к ее представлению в базах данных (БД), и поэтому процесс преобразования ИТВ в формат БД можно формализовать. Проектирование реляционной таблицы (РТ) на основе использования существующих инструментальных средств рассмотрено в работах [3-7]. В предлагаемой статье рассматриваются методика и средства, которые позволят выявить и исключить отличия ИТВ от РТ и выполнить автоматизированное преобразование заполненных ИТВ в РТ.

1. Основные различия между информацией, представленной в табличном виде, и информацией, представленной в виде реляционной таблицы

Примерами ИТВ могут быть ведомости, прайс-листы, словари, списки и многое другое. Главная задача разработчиков такого рода таблиц – это упрощение восприятия информации пользователем.

В работах [1, 2] рассмотрено понятие ИТВ, определены следующие свойства ИТВ, отличающие ее от РТ:

- информация интуитивно воспринимается ее потребителями как таблицы;
- нередко отсутствуют разделители строк и разделители столбцов;
- элементы данных нередко размещаются в нескольких строках;
- типы элементов данных внутри одного столбца могут различаться;
- заголовки могут включать в себя подзаголовки;
- заголовки и/или подзаголовки одноименного столбца нередко размещаются в нескольких строках.

Появляющиеся новые версии электронных таблиц, текстовых редакторов, текстовых процессоров, HTML средств, систем обработки статистической информации и другие имеют наряду с перечисленными выше дополнительные свойства:

- отражение семантики данных посредством цвета, фона, шрифта и т.п.;
- повторное использование заголовков и подзаголовков;
- отражение заголовков и подзаголовков посредством месторасположения в таблице, цвета, фона, шрифта и т.п.;
- использование значений доменов атрибутов в качестве подзаголовков;
- использование комбинированных подзаголовков;
- использование одноименных доменов;
- отсутствие первичных ключей;
- отсутствие внешних ключей.

Для того чтобы учесть эти свойства, введем понятие расширенная информация табличного вида (ИТВР)

Формат ИТВР может быть бумажный и электронный, формат текстовых редакторов, текстовых процессоров, электронных таблиц и др. В связи с этим возникает вторая проблема – проблема преобразования форматов.

С актуальностью проблемы преобразования ИТВР в РТ авторы столкнулись в реальных разработках, обнаружив что, с одной стороны, реляционные БД исключительно редко создаются на пустом месте (чаще всего имеются значительные объемы ИТВР), с другой стороны, накопилось множество ИТВР, которое необходимо обрабатывать средствами современных систем управления базами данных (СУБД).

К сожалению, в настоящее время нет теоретических и практических разработок, которые могли бы в полном объеме решить проблемы преобразования ИТВР в формат реляционных БД. Как сказал один из основоположников теории проектирования БД Дейт К. Дж.: “Проектирование БД – это скорее искусство, чем наука” [3]. А преобразование ИТВР в

формат реляционных БД – это важнейшая проблема, стоящая перед проектировщиками БД.

В рамках методики преобразования ИТВР в РТР необходимо задействовать алгоритмы:

- алгоритм исключения заголовков в содержимом таблиц;
- алгоритм исключения подзаголовков в содержимом таблиц;
- алгоритм исключения комбинированных подзаголовков;
- алгоритм исключения пустых подзаголовков;
- алгоритм избавления от вертикальных подзаголовков;
- алгоритм приведения значений атрибутов таблиц к одному типу;
- алгоритм исключения дублирования записей в таблицах;
- алгоритм исключения всех возможных сочетаний подзаголовков (внешних, внутренних, вертикальных);
- алгоритм исключения пустых заголовков;
- алгоритм исключения всех видов вертикальных подзаголовков;
- алгоритм исключения пустых записей;
- алгоритм приведения значений доменов к одному типу.

Для назначения ключевых полей в преобразованных таблицах ИТВР необходимо задействовать следующие алгоритмы:

- алгоритм выявления столбцов с уникальными значениями его элементов;
- алгоритм выявления сочетания столбцов с уникальными сочетаниями соответствующих им элементов;
- алгоритм поиска минимальных первичных ключевых полей, которые включают в себя один атрибут;
- алгоритм поиска минимальных первичных ключевых полей, которые включают в себя несколько атрибутов;
- алгоритм выявления атрибутов, которые входят в первичный ключ и содержат уникальные значения;
- алгоритм выявления внешних ключевых полей.

Обобщая перечисленные алгоритмы, можно назвать основные отличия моделей ИТВР от модели РТ:

- 1) ИТВР, как правило, представлены таблицами, которые не удовлетворяют требованиям к РТ. (В РТ недопустимы перечисленные выше свойства ИТВР).
- 2) В ИТВР первичные и внешние ключи не определены.
- 3) Между таблицами ИТВР нет связей.

2. Разработка операторной модели преобразования ИТВР в РТ

В мероприятиях разработки операторной модели за основу взят подход, который предложен в работе [2].

Введем понятие "Расширенная реляционная таблица" (РТР). Эта таблица, в которой недопустимы особенности ИТВР, перечисленные выше.

Укрупненная модель процесса формирования РТР на основе использования заполненных ИТВР представлена на рисунке 1,а. Использование в модели изображения человека подчеркивает то, что разрабатывается модель человеко-машинного процесса.

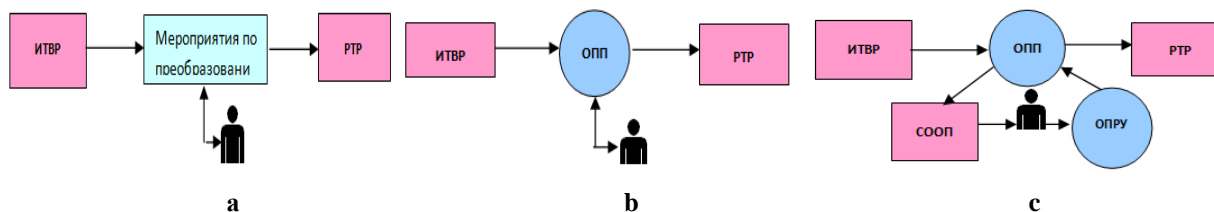


Рис. 1. Модель процесса формирования РТР на основе использования заполненных ИТВР

На рисунке для наглядности операторы и состояния модели выделены разными цветами.

Активное и определяющее участие разработчика в процессе преобразования ИТВР в РТР позволяет в полной мере задействовать его творческий потенциал. В этой связи стоит задача разработки следующих средств:

- средств, которые позволяют активно участвовать разработчику в процессе преобразования ИТВР в РТР;
- средств, которые позволяют использовать существующие системы проектирования реляционных БД.

Введем оператор преобразования ИТВР в РТР. Результатом выполнения этого оператора является объект, соответствующий предложенной ранее модели РТР. Операндами оператора ОПП, как видно из рисунка 1,б, являются модель ИТВР и модель РТР.

Таким образом: $РТР = ОПП(ИТВР)$.

Разработчик РТР должен принимать решение по результатам выполнения ОПП. В связи с этим необходима система оценок результатов выполнения ОПП. Назовем ее СООП. Для реализации указаний разработчика необходимы соответствующие средства. Оператор, соответствующий этим средствам, назовем ОПРУ.

В свете высказанных соображений укрупненная операторная модель формирования РТР на основе использования заполненных ИТВР примет вид (рис. 1,с).

В целях использования типовых компонентов модели исключим изображение человека, определив тем самым, что разработчик задействован в СООП и ОПРУ. В результате получим следующее представление модели формирования РТР на основе использования заполненных ИТВР (рис. 2,а).

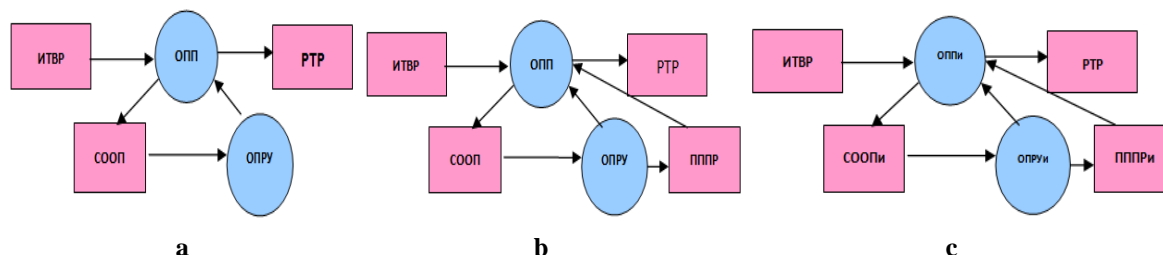


Рис. 2. Компоненты модели процесса формирования РТР на основе использования заполненных ИТВР

Следует отметить, что, несмотря на внешнее сходство рис. 2,b и 2,с, в них задействованы различные состояния и оператор и они иллюстрируют различные процессы.

Для выполнения ОПП и ОПРУ необходима система оценок. Эта система должна обеспечить оценку результатов указаний разработчика (ОПРУ) и обеспечить переход к оператору преобразования ОПП. Назовем соответствующее состояние модели ПППР (подсистема преобразований).

Пятая операторная форма модели автоматизированного преобразования ИТВР в РТР представлена на рис. 2,b.

Используя промежуточную операторную модель автоматизированного преобразования ИТВР в РТР, мы можем получить в операторной форме следующие выражение:

$\text{РТР} = \text{ОПП}(\text{ИТВР}, \text{СООП}, \text{ПППР});$

$\text{ППР} = \text{ОПРУ}(\text{СООП}).$

Построенная модель является предельным упрощением описания реального процесса преобразования информации ИТВР к РТР.

В полной операторной модели методики преобразования необходимо рассмотреть все формы представления ИТВР. При этом необходимо рассмотреть как самый благоприятный случай представления ИТВР для преобразования в РТР, так и все неблагоприятные случаи.

В качестве компонента модели процесса преобразования следует включить манипуляции, связанные с импортом таблиц, которые представлены не в формате инструментальной СУБД. Для этого введен соответствующий фрагмент, который отражен на рис. 2,с.

Здесь:

ОППи – оператор импорта таблиц ИТВР;

СООПи – состояние модели, соответствующее автоматической генерации системы оценок результатов импорта;

ОПРУи – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов импорта;

ПППРи – состояние модели, соответствующее формированию системы оценок результатов импорта разработчиком РТ;

Даже в самом благоприятном случае остается необходимость назначения первичных и внешних ключей, формирования связей между таблицами. В ИТВР их просто нет.

В реальных ситуациях ИТВР, как правило, не удовлетворяет требованиям к реляционным таблицам. Методика должна обеспечить решение задачи для всех случаев.

Вспомогательная таблица возможных ситуаций представления ИТВР

Для полноты картины приведем таблицу, в которой представлены возможные ситуации представления ИТВР (таблица 1).

Таблица 1. Возможные ситуации представления ИТВР

№	НРЛ	НРЛвп	НРЛвс	НКЛ	НКЛт	НКЛв	НСВ
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1
...
20	0	0	1	0	0	1	1
21	0	0	1	0	1	0	0
22	0	0	1	0	1	0	1
23	0	0	1	0	1	1	0
24	0	0	1	0	1	1	1
25	0	0	1	1	0	0	0
26	0	0	1	1	0	0	1
27	0	0	1	1	0	1	0
28	0	0	1	1	0	1	1
29	0	0	1	1	1	0	0
30	0	0	1	1	1	0	1
31	0	0	1	1	1	1	0
32	0	0	1	1	1	1	1
33	0	1	0	0	0	0	0
34	0	1	0	0	0	0	1
35	0	1	0	0	0	1	0
36	0	1	0	0	0	1	1
37	0	1	0	0	1	0	0
38	0	1	0	0	1	0	1
39	0	1	0	0	1	1	0
40	0	1	0	0	1	1	1
...
5040	1	1	1	1	1	1	1

В таблице представлены все возможные сочетания вариантов несоответствия ИТВР таблицам РТР. Цифрой 1 обозначено наличие соответствующего варианта, цифрой 0 - отсутствие. В принципе необходимо рассмотреть все сочетания.

Здесь столбцы имеют следующие значения:

НРЛ – таблицы ИТВР не удовлетворяют традиционным требованиям к реляционным таблицам;

НКЛ – в таблицах ИТВР отсутствуют данные о ключевых полях;

НСВ – в таблицах ИТВР отсутствуют данные о связях между таблицами.

На рис 2,с рассмотрен фрагмент модели для случая 1, а именно случай, когда таблицы ИТВР удовлетворяют требованиям к реляционным таблицам, в таблицах ИТВР имеются данные о ключевых полях, в таблицах ИТВР имеются данные о связях между таблицами;

НРЛвп – в таблицах ИТВ присутствуют внутренние подзаголовки;

НРЛвс – в таблицах ИТВ заголовки размещены в первом столбце;

НКЛт – для формирования внешнего ключа в таблицах необходимо более 3-х атрибутов;

НКЛв – в таблицах не назначены внешние ключи.

Пример 1 использования модели: *простой импорт ИВТР в РТР.*

Нетрудно предположить, что самым благоприятным случаем представления ИТВР для их автоматизированного преобразования в РТР является такой случай, когда ИТВР удовлетворяют требованиям к реляционным таблицам. Тогда значительная часть работ по преобразованию сводится к простому импорту таблиц, который обеспечивается во всех инструментальных СУБД. Поэтому в модель процесса следует включить манипуляции, связанные с импортом таблиц, которые представлены не в формате инструментальной

СУБД или в формате другой инструментальной СУБД. Для этого необходимо в модель ввести соответствующий фрагмент, который отражен на рис. 3

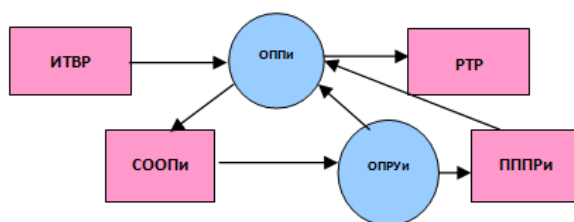


Рис. 3. Пример использования операторной модели: импорта ИТВР в РТР

Обозначения: ОППи – оператор импорта таблиц ИТВР;

СООПи – состояние модели, соответствующее автоматической генерации системы оценок результатов импорта;

ОПРУи – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов импорта;

ПППРи – состояние модели, соответствующее формированию системы оценок результатов импорта разработчиком РТ.

На рис. 3 рассмотрен фрагмент модели для случая, когда таблицы ИТВР удовлетворяют требованиям к реляционным таблицам, в таблицах ИТВР имеются данные о ключевых полях, в таблицах ИТВР имеются данные о связях между таблицами. Этот пример соответствует 1-му сочетанию в таблице 1.

Пример 2 использования модели: *ключевые поля в ИТВР не определены.*

Этот пример соответствует 9-му сочетанию в таблице 1.

В данном случае ключевые поля не определены, таблицы ИТВР удовлетворяют всем перечисленным требованиям к реляционным таблицам, в таблицах ИТВР имеются данные о связях между таблицами. Соответствующая операторная модель выглядит (рис. 4).

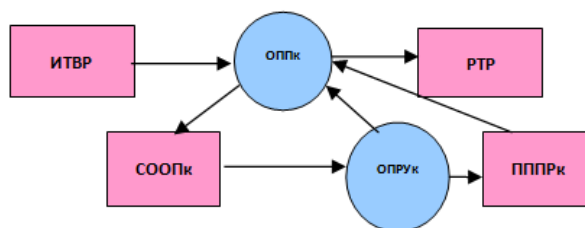


Рис. 4. Операторная модель процесса автоматизированного назначения ключевых полей в заполненных ИТВР

На рисунке 4 отражены новые операторы и состояния модели. В частности, здесь:

ОППк – оператор назначения ключевых полей в ИТВР;

СООПк – состояние модели, соответствующее автоматической генерации системы оценок результатов назначения ключевых полей в ИТВР;

ОПРУк – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов назначения ключевых полей в ИТВР;

ПППРк – состояние модели, соответствующее формированию системы оценок результатов назначения ключевых полей в ИТВР разработчиком РТР.

Пример 3 использования модели: *внешние ключевые поля не определены.*

Этот пример соответствует 3-му сочетанию в таблице 1.

Внешние ключевые поля – это поля подчиненных таблиц, которые связаны с первичными ключами основных таблиц.

В следующей итерации построения модели процесса необходимо отразить средства назначения ключевых полей на основе 3-х и более атрибутов и средства назначения внешних ключевых полей. Отразим фрагмент расширенной операторной модели процесса назначения ключевых полей на рис. 5.

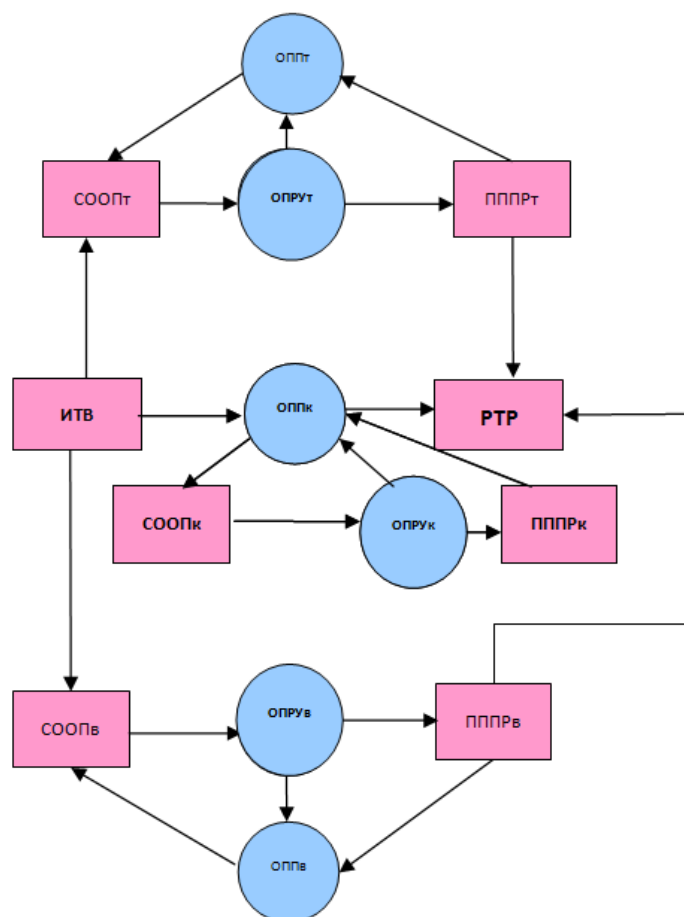


Рис. 5. Фрагмент расширенной операторной модели процесса назначения ключевых полей.

Здесь:

ОППт – оператор назначения ключевых полей с числом атрибутов более 3-х в ИТВР;

СООПт – состояние модели, соответствующее автоматической генерации системы оценок результатов назначения ключевых полей с числом атрибутов более 3-х в ИТВР;

ОПУт – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов назначения ключевых полей с числом атрибутов более 3-х в ИТВР;

ПППт – состояние модели, соответствующее формированию системы оценок результатов назначения ключевых полей с числом атрибутов более 3-х в ИТВР разработчиком РТР;

ОППв – оператор назначения внешних ключевых полей в ИТВР;

СООПв – состояние модели, соответствующее автоматической генерации системы оценок результатов назначения внешних ключевых полей в ИТВР;

ОПРУв – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов назначения внешних ключевых полей в ИТВР;

ПППРв – состояние модели, соответствующее формированию системы оценок результатов назначения внешних ключевых полей в ИТВР разработчиком РТР.

Пример 4 использования модели: *структура таблиц ИТВР не соответствует требуемой структуре таблиц РТ.*

Рассмотрим еще один характерный случай (позиция 16 в таблице 1), а именно случай, когда структура таблиц ИТВР не соответствует требуемой структуре таблиц РТ (эти несоответствия перечислены в начале статьи). При этом в таблицах ИТВР имеются данные о связях между таблицами, в таблицах ИТВР имеются все ранее перечисленные сведения о ключевых полях.

Соответствующая операторная модель выглядит следующим образом (рис. 6).

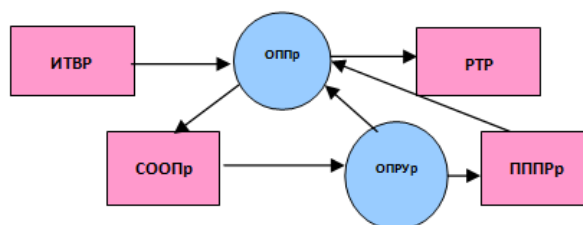


Рис. 6. Операторная модель процесса автоматизированного преобразования ИТВР в РТР

Здесь:

ОППр – оператор преобразования ИТВР в РТР;

СООПр – состояние модели, соответствующее автоматической генерации системы оценок результатов преобразования ИТВР в РТР;

ОПУр – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов преобразования ИТВР в РТР;

ПППРр – состояние модели, соответствующее формированию системы оценок результатов преобразования ИТВР в РТР разработчиком РТР.

В этом фрагменте модели отражена необходимость разработки следующих алгоритмов:

- алгоритма исключения внутренних подзаголовков в ИТВР;
- алгоритма исключения заголовков, сформированных в первом столбце таблицы ИТВР;
- алгоритма исключения сложных подзаголовков в ИТВР;
- алгоритма исключения повторяющихся записей в ИТВР.

Последний алгоритм тривиален и реализуется средствами инструментальной СУБД. Предпоследний алгоритм является комбинированием двух предыдущих алгоритмов. По-

этому в расширенной операторной модели преобразования ИТВР в РТР отобразим необходимость разработки только первых двух алгоритмов (рис. 7).

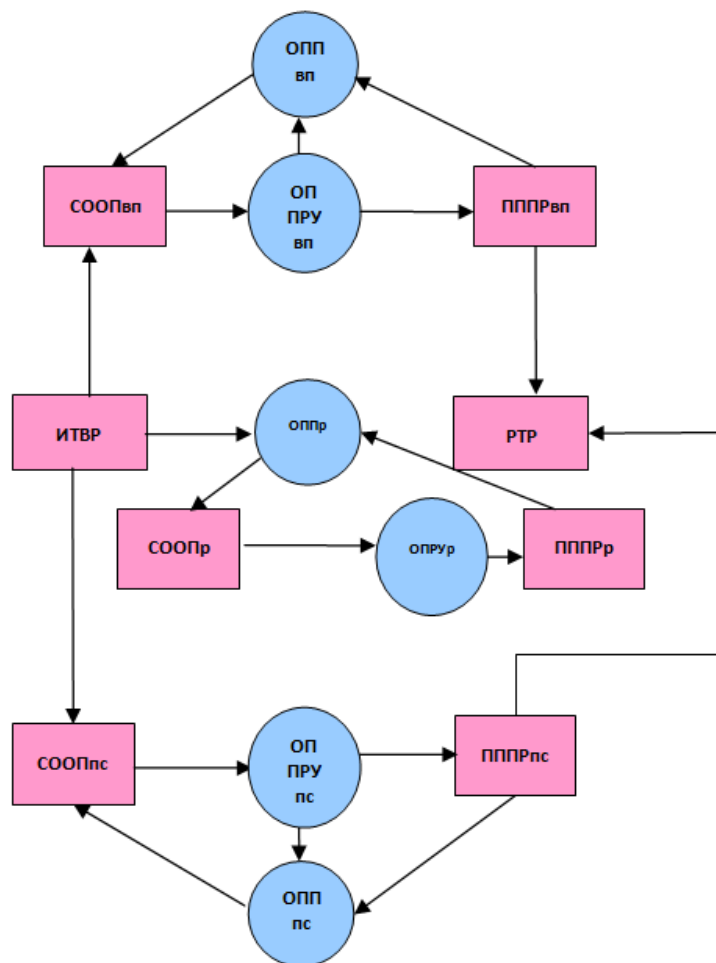


Рис. 7. Расширенная операторная модель преобразования ИТВР в РТР

Здесь:

ОПП_{вп} – оператор исключения внутренних подзаголовков в ИТВР;

СООП_{вп} – состояние модели, соответствующее автоматической генерации системы оценок результатов исключения внутренних подзаголовков в ИТВР;

ОПРУ_{вп} – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов исключения внутренних подзаголовков в ИТВР;

ПППР_{вп} – состояние модели, соответствующее формированию системы оценок результатов исключения внутренних подзаголовков в ИТВР разработчиком РТР;

ОПП_{пс} – оператор исключения заголовков первого столбца в ИТВР;

СООП_{пс} – состояние модели, соответствующее автоматической генерации системы оценок результатов исключения заголовков первого столбца в ИТВР;

ОПРУ_{пс} – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов исключения заголовков первого столбца в ИТВР;

ПППР_{пс} – состояние модели, соответствующее формированию системы оценок результатов исключения заголовков первого столбца в ИТВР разработчиком РТР;

Строго говоря, необходимо рассмотреть все 5040 случаев и в соответствии с каждым случаем детализировать операторную модель методики. Выполненные выкладки для всех случаев типовые. В связи с этим представим результирующую операторную модель процесса преобразования ИТВР в РТР, которая соответствует случаю 5040 (рис. 8). Кроме того, последнее сочетание перекрывает все возможные случаи.

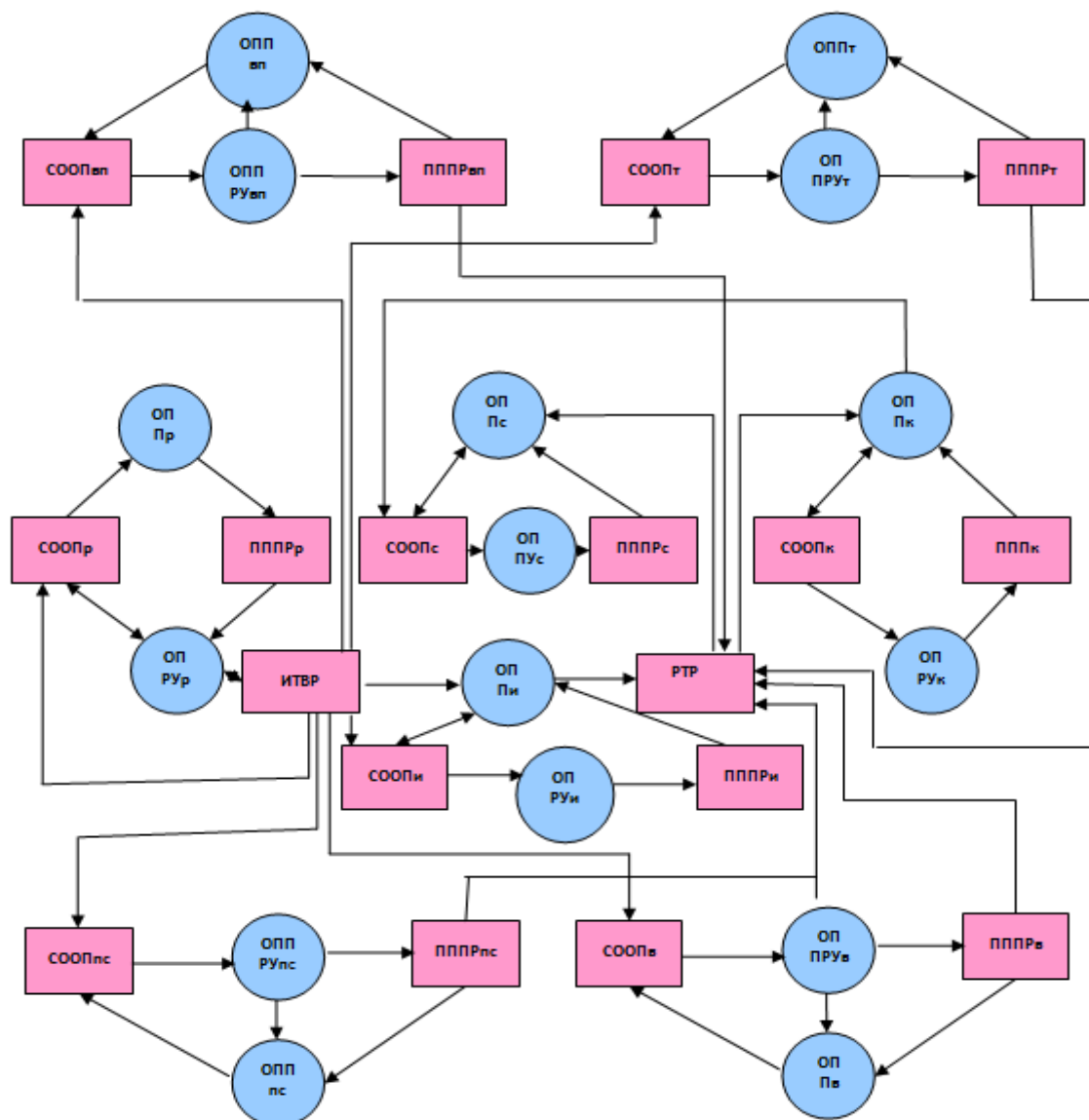


Рис. 8. Результирующая операторная модель процесса преобразования ИТВР в РТР

Здесь:

ОППс – оператор формирования связей между таблицами ИТВ;

СООПс – состояние модели, соответствующее автоматической генерации системы оценок результатов импорта;

ОПРУс – оператор принятия решений относительно управляющих воздействий после анализа системы оценок результатов формирования связей;

ПППРс – состояние модели, соответствующее формированию системы оценок результатов формирования связей разработчиком РТ.

При построении результирующей операторной модели процесса преобразования ИТВР в РТР использованы фрагменты модели, представленные на рисунках 4 - 7. Принято во внимание то, что модели ИТВР и РТР для всех случаев едины. Принято во внимание то, что модель, представленная на рис. 8, отражает все возможные сочетания.

Эта операторная модель соответствует методике автоматизированного формирования РТР на основе использования заполненных ИТВР. Она позволяет определить состав процедур, которые необходимо реализовать для создания полноценной программной подсистемы, может быть использована в качестве формализованного руководства пользователя и выступает в качестве источника для дальнейшей формализации процесса преобразования и его исследования.

Модель процесса автоматизированного преобразования ИТВР в РТ (рис. 8) позволяет назвать основные процедуры подсистемы, которые необходимо реализовать. Этим процедурам соответствуют следующие кортежи операторов:

ОПП = {ОППи, ОППр, ОППк, ОППс, ОППт, ОППв, ОППвп, ОППпс}.

ОПРУ = {ОПРУи, ОПРУр, ОПРУк, ОПРУс, ОПРУт, ОПРУв, ОПРУвп, ОПРУпс}.

ПППР = {ПППРи, ПППРр, ПППРк, ПППРс, ПППРп, ПППРв, ПППРвп, ПППРпс}.

СООП = {СООПи, СООПр, СООПк, СООПс, СООПп, СООПв, СООПвп, СООПпс}.

Заключение

Состав и назначение перечисленных операторов, состав и описание состояний модели, взаимосвязи между ними позволяют, с одной стороны, сформулировать требования к инструментальным средствам, которые необходимо реализовать для обеспечения автоматизированного преобразования ИТВР в РТР, а с другой - определяют методику преобразования.

Несмотря на достоинства построенной модели, необходимо ее формальное исследование и уточнение, чему и посвящены дальнейшие работы авторов.

Список литературы

- [1]. Брешенков А.В. Неформальная постановка проблемы преобразования информации табличного вида в файлы баз данных // Сб. трудов АУ МВД России "Актуальные вопросы технологий в деятельности органов внутренних дел". М., 2004. С. 55– 70.
- [2]. Брешенков А.В. Методология проектирования реляционных баз данных с использованием данных табличного вида: дис. ... доктора техн. наук. М., 2007. 417 с.
- [3]. Дейт К. Дж. Введение в системы баз данных. 8-е изд., доп.: Пер. с англ. М.: Вильямс. 2005. 1328 с. [Date C.J. An Introduction to Database Systems, 8th Edition]
- [4]. Date C.J. Date on Database: Writings 2000-2006. Part 5: Further Relational Misconceptions. Chapter 19: There's Only One Relational Model. Apress. 2006. 568 p. P. 345-361. (Contents: <http://123doc.org/document/1345656-date-on-database.htm?pageh=8>)

- [5]. Codd E.F. The Relational Model For Database Management Version 2. Inc. Boston, MA, USA: Addison – Wesley. 1990. 511 p.
- [6]. Хансен Г, Хансен Дж. Базы данных. Разработка и управление. М.: Бином. 2001. 704 с.
- [7]. Ульман Д., Уидом Д. Введение в системы баз данных. М.: Вильямс. 2003. 1088 с.