

СТАРООСКОЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ  
ИНСТИТУТ им. А.А. УГАРОВА



**XI**

**ВСЕРОССИЙСКАЯ  
НАУЧНО-ПРАКТИЧЕСКАЯ  
КОНФЕРЕНЦИЯ**  
С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

**«Современные проблемы горно-металлургического комплекса.  
Наука и производство»**

**I ТОМ**

**3–5 декабря 2014 г.  
г. Старый Оскол**

Министерство образования и науки Российской Федерации  
Старооскольский технологический институт им.А.А. Угарова  
(филиал) федерального государственного автономного образовательного  
учреждения  
высшего профессионального образования  
«Национальный исследовательский технологический университет «МИСиС»

**Современные проблемы  
горно-металлургического комплекса.  
Наука и производство**

Материалы Одиннадцатой Всероссийской научно-практической  
конференции, с международным участием

3-5 декабря 2014г.

г.Старый Оскол

Старый Оскол

2014

# РЕИНЖИНИРИНГ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА ОСНОВЕ ПОИСКА И АНАЛИЗА ФУНКЦИОНАЛЬНО-ЗАВИСИМЫХ АТТРИБУТОВ

Филатов В.А., Чапланова Е.Б., Радченко В.А.

Харьковский национальный университет радиоэлектроники, Харьков, Украина

[filatov\\_val@ukr.net](mailto:filatov_val@ukr.net), [chaplanova@gmail.com](mailto:chaplanova@gmail.com), [cayrad@gmail.com](mailto:cayrad@gmail.com)

*Аннотация.* Рассмотрена задача реинжиниринга информационной системы на основе реляционной базы данных. Предложен подход к определению функционально-зависимых атрибутов базы данных на этапе реинжиниринга и модифицированный алгоритм синтеза логической схемы реляционной базы данных.

*Ключевые слова:* реинжиниринг, информационная система, реляционная база данных, отношение, схема базы данных, функциональные зависимости, нормализация.

## REENGINEERING RELATIONAL DATABASE INFORMATION SYSTEM BASED ON SEARCH AND ANALYSIS FUNCTIONAL THE DEPENDENT ATTRIBUTE

Filatov V.A., Chaplanova E.B., Radchenko V.A.

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

[filatov\\_val@ukr.net](mailto:filatov_val@ukr.net), [chaplanova@gmail.com](mailto:chaplanova@gmail.com), [cayrad@gmail.com](mailto:cayrad@gmail.com)

*Abstract.* In this paper we solve the problem of reengineering the information system based on a relational database. We assume the new approach to determine function-dependent attributes of the database during reengineering and provide modified algorithm to synthesize database logical schema.

*Keywords:* reengineering, information systems, relational database, relation, the database schema, functional dependencies, normalization.

### Введение

Реинжиниринг информационных систем, основой которых являются базы данных (БД), в последнее время привлекают все большее внимание специалистов в области современных информационных технологий, и это обусловлено рядом причин. В процессе своего существования БД неизбежно претерпевает изменения, связанные с непостоянством предметной области (например, усовершенствование бизнес-процессов, возникновение новых требований со стороны пользователей). Ключевым компонентом, во многом

определяющем характеристики функционирования реляционной базы данных (РБД), является логическая схема, представляющая собой описание таблиц данных и связей между ними.

На сегодняшний день наибольшее распространение получили реляционные базы данных, обеспечивающие наилучшее сочетание надежности, простоты использования и производительности для решения различных задач. Поэтому выполнение задач сопровождения зачастую невозможно осуществить без внесения изменений в логическую схему БД. При этом необходимо учитывать существующие взаимосвязи между данными, заданные в процессе первоначального проектирования и продиктованные предметной областью. От того, насколько квалифицировано спроектирована БД, зависит эффективность функционирования информационной системы (ИС) в целом и возможности адаптации под изменяющиеся требования.

### **1. Цель проводимых исследований**

Разработать и исследовать метод реинжиниринга реляционных баз данных на основе поиска, анализа и классификации множества функциональных зависимостей в текущем наборе данных. Множество функциональных зависимостей предполагается использовать в качестве входных данных для метода синтеза схемы реляционной базы данных в третьей нормальной форме.

### **2. Основная часть**

В общем случае реинжиниринг включает в себя два этапа: формирование структуры РБД, отвечающей новым требованиям и перенос данных в эту структуру.

В методе реинжиниринга схемы данных предлагается подход с использованием множества строгих функциональных зависимостей. Такой подход позволяет перейти от исходного отношения к набору отношений в третьей нормальной форме путем применения последовательности правил декомпозиции.

Представим исходную РБД как множество  $DB = \{\rho_1, \dots, \rho_n\}$ , где  $\rho_i$  – отношение БД,  $i = \overline{1, n}$  ( $n$  – количество отношений БД). Выразим  $\rho_i$  через набор  $\langle \sigma_i, p_i \rangle$ , где  $\sigma_i$  – логическая схема отношения, а  $p_i$  – экземпляр отношения (множество кортежей). Логической схемой отношения будем считать  $\sigma_i = \langle R_i, F_i \rangle$ , где  $R_i$  – множество атрибутов, а  $F_i$  – множество функциональных зависимостей (ФЗ), выполняющихся на  $R_i$ .

С другой стороны,  $DB = \langle \Sigma, P \rangle$ , где  $\Sigma = \{\sigma_1, \dots, \sigma_n\}$  – логическая схема, а  $P = \{p_1, \dots, p_n\}$  – экземпляр данных РБД. Также логическую схему  $\Sigma$  можно выразить как набор  $\Sigma = \{R, F\}$ , где  $R = \bigcup R_i$  – общее множество атрибутов, а  $F = \bigcup F_i$  – общее множество ФЗ.

В процессе нормализации при проектировании исходной структуры РБД происходит декомпозиция универсального отношения  $U$  согласно множеству ФЗ, которые выполняются

на  $U$ . Целью нормализации является устранение избыточности данных, а также аномалий обновления/удаления. Как правило, достаточной считается нормальная форма Бойса-Кодда (НФБК), в которой отсутствуют какие-либо аномалии. Но ввиду того, что не все схемы можно привести к НФБК, основной принято полагать третью нормальную форму (3НФ) [1]. Обозначим через  $\Sigma'$  текущую логическую схему РБД, которая является результатом внесения ряда изменений согласно изменениям требований к ИС. Текущее множество атрибутов обозначим как  $R'$ . В процессе функционирования РБД также происходит переход от  $R$  к  $R'$ , где  $R'$  – текущий экземпляр данных.

Нельзя гарантировать, что  $\Sigma'$  находится в 3НФ: если  $R' \neq R$ , то из этого следует  $F' \neq F$ , где  $F'$  – текущее множество ФЗ. Поскольку имеет место  $\Sigma' = \langle R', F' \rangle$ , то необходимой является проверка текущей логической схемы  $\Sigma'$  на соответствие 3НФ, а также приведение к ней в противном случае.

Для решения поставленной задачи предлагается построить 3НФ для текущей логической схемы, используя предложенный Ф. Бернштейном метод синтеза [2]. Он принимает в качестве входного параметра множество ФЗ и формирует множество возможных реализаций схем  $S = \{\hat{\Sigma}_1 \dots \hat{\Sigma}_m\}$ , которые находятся в 3НФ. Следует отметить, что данная задача не является тривиальной. Множество  $F'$ , которое является исходными данными для метода синтеза, не полностью определено ввиду того, что при изменении множества атрибутов изменяется и множество ФЗ, соблюдающихся на этих атрибутах. Поэтому первым этапом является нахождение множества  $F'$ , справедливого для  $R'$ .

Пусть  $F_S$  – множество тех ФЗ, которые представляется возможным получить путем анализа ограничений целостности реляционной СУБД, таких как первичные и внешние ключи. Обозначим с помощью  $F_H$  то множество ФЗ, которое представляет собой неявные зависимости, о существовании которых не было известно во время первоначального проектирования. Они существуют в виде закономерностей в данных, установившихся в процессе функционирования ИС, и будут использованы в процессе синтеза целевой схемы. Таким образом, искомое множество ФЗ  $F' = F_S \cup F_H$ ; нахождение  $F_S$  не рассматривается в данной работе.

Для определения множества  $F_H$  существует ряд методов выявления зависимостей из экземпляров РБД. В основном, они используются в задачах интеллектуального анализа данных и позволяют находить приблизительные ФЗ (ПФЗ). Отличие от классических «строгих» ФЗ состоит в том, что допускается существование ФЗ, даже если в РБД присутствуют строки, нарушающие корректность ФЗ. Это допущение опирается на предположение о том, что в процессе функционирования РБД возможно внесение «ошибочных» строк. Под «ошибочной» строкой следует понимать такой кортеж, который не

противоречит существующим ограничениям целостности, но нарушает соблюдавшуюся до этого момента неявную ФЗ. Одним из таких методов является метод Tane [3] и его модификации, послуживший основой для других подобных решений. Известна классическая модификация данного метода, позволяющая выявлять только строгие ФЗ [4]. В качестве входных данных необходимо предоставить экземпляр  $P'$ ; результатом работы метода является искомое множество  $F_H$ . Тем не менее, корректность результата не гарантируется для данных, содержащих пустые значения (NULL). Поддержка пустых значений в задаче выявления ФЗ не рассматривается в данной работе.

Исходными данными для решения подзадачи нахождения множества  $F_H$  являются: логическая схема реляционной БД  $\Sigma = \{\sigma_i, i = \overline{1, n}\}$ , где  $\sigma_i$  – схема одного отношения, входящего в БД,  $n$  – количество отношений; схема отношения  $\sigma_i = \langle R_i, F_i \rangle$ , где  $R_i$  – носитель отношения (множество атрибутов), а  $F_i$  – множество функциональных зависимостей (ФЗ), удовлетворяющих данному отношению.  $P = \{\rho_i, i = \overline{1, n}\}$  – множество отношений рассматриваемой БД [5]. Основные этапы метода синтеза:

1. Устранение лишних атрибутов. Пусть  $F$  – исходный набор ФЗ. После устранения лишних атрибутов из левых частей каждой ФЗ в  $F$  в результате будет получено множество  $F'$ . Атрибут является лишним, если его устранение не влияет на замыкание множества ФЗ.

2. Нахождение покрытия. Необходимо найти такое множество  $H$  для  $F'$ , что любую ФЗ из  $F'$  можно было вывести, используя зависимости из  $H$ .

3. Разбиение. Разбить  $H$  на группы, такие, что все ФЗ в каждой группе будут иметь одинаковые левые части.

4. Объединение эквивалентных ключей. Пусть  $J = \emptyset$ . Для каждой пары групп  $H_i$  и  $H_j$  с левыми частями  $X$  и  $Y$ , соответственно, необходимо объединить  $H_i$  и  $H_j$  если существует биекция  $X \leftrightarrow Y$  в  $H^+$ . Для каждой такой биекции добавляем  $X \rightarrow Y$  и  $Y \leftarrow X$  в  $J$ . Проводится проверка для каждого атрибута  $A \in Y$ : если  $X \rightarrow A$  находится в  $H$ , необходимо удалить его из  $H$ . То же самое проделывается для каждого  $X \rightarrow B$  в  $H$  с  $B \in X$ .

5. Устранение транзитивных зависимостей. Для этого необходимо найти  $H' \subseteq H$  такое, что  $(H' + J)^+ = (H + J)^+$  и ни одно подмножество, принадлежащее  $H'$ , не должно обладать этим свойством. Затем нужно добавить каждую ФЗ, принадлежащую  $J$ , в соответствующую группу  $H'$ .

6. Построение отношения. Для каждой группы построить отношение, состоящее из всех атрибутов, находящихся в такой группе. Каждое множество атрибутов, которое находится в левой части любой ФЗ в группе, является ключом отношения. (Шаг 1 гарантирует, что такое множество не будет содержать лишних атрибутов). Все ключи,

найденные таким образом, будут называться синтезированными. Множество построенных отношений составляет схему для заданного множества ФЗ.

Минимальность данного метода гарантируется тем, что все избыточные покрытия дают в результате одинаковое количество отношений, поскольку количество классов эквивалентности синтезированных ключей является одинаковым для всех избыточных покрытий некоторого множества ФЗ [6]. В случае, если  $\Sigma' \notin S$ , текущая схема не находится в ЗНФ, и для получения схемы в ЗНФ наиболее очевидным вариантом является выбор одной из реализаций  $\hat{\Sigma} \in S$  в качестве целевой. В случае, когда реализаций  $\hat{\Sigma}$  несколько, критерием выбора предлагается использование экспертной оценки, поскольку они все являются корректными с позиции условий ЗНФ.

Предложенный подход к задаче реинжиниринга реляционной базы данных позволяет осуществить синтез схемы данных, а также получить оценки соответствия логической схемы третьей нормальной форме. В конечной структуре будут учтены не только явные взаимосвязи между данными, выраженные с помощью ограничений целостности реляционной системы, но и скрытые зависимости предметной области, которые установились исторически в процессе функционирования ИС.

### **Заключение**

В статье рассмотрен подход к задаче реинжиниринга информационных систем на основе реляционных баз данных. Выделены проблемы реинжиниринга РБД, исследован этап формирования логической схемы базы данных, являющийся общим для задач адаптации и рефакторинга. Рассмотрена подзадача проверки нахождения логической схемы РБД в ЗНФ, показано, что ее решение сопряжено с рядом трудностей, в частности, необходимостью нахождения множества функциональных зависимостей. Предложен подход для нахождения функциональных зависимостей из экземпляров данных РБД.

Предложенный метод поиска функциональных зависимостей дает возможность включить в процесс реинжиниринга те взаимосвязи предметной области, которые не были включены на начальном этапе проектирования. Это позволяет гарантировать оптимальность конечной логической схемы, полученной с помощью метода синтеза, и, как следствие – минимизацию избыточности хранимых данных в процессе функционирования реляционной базы данных.

### **Список литературы**

1. Bernstein P.A. Synthesizing Third Normal Form Relations from Functional Dependencies / ACM Transactions on Database Systems Volume 1 Issue 4. – 1976. – С. 277 – 298.
2. Nick Rossiter. Re-engineering relational databases: the way forward: ISWSA '11, ACM New York, NY, USA, 2011. – 17 с.

3. Henrard, J. Data dependency elicitation in database reverse engineering: Software Maintenance and Reengineering Conference, 2001. – С. 11-19.
4. Мейер Д. Теория реляционных баз данных // Пер. с англ. – М.: Мир, 1987. – 609 с.
5. В.А. Радченко. Модификация метода выявления функциональных зависимостей в реляционных базах данных / Інформаційні технології в навігації і управлінні: стан та перспективи розвитку. Матеріали другої міжнародної науково-технічної конференції. – К.: ДП «ЦНДІ НіУ», 2011. – 52 с.
6. В.А. Радченко, С.С. Танянский. Выявление скрытых зависимостей между данными в задачах реинжиниринга информационных систем / Системи обробки інформації. Випуск 3(101) – Х.: ХУПС, 2012. – 268 с.

УДК 004.93'1, 65.015.2

## **КОМПЬЮТЕРНОЕ ЗРЕНИЕ КАК ОСНОВА АВТОМАТИЗАЦИИ УЧЕТА РАБОЧЕГО ВРЕМЕНИ НА ПРОИЗВОДСТВЕ**

**М.А. Цуканов, О.П. Ульянова**

Старооскольский технологический институт им А.А. Угарова (филиал) ФГАОУ ВПО  
"Национальный исследовательский технологический университет "МИСиС", Старый Оскол  
tsukanov\_m\_a@mail.ru, ulianov032223@mail.ru

*Аннотация.* В данной статье рассмотрена общая классификация методов компьютерного зрения, предложен автоматизированный подход к нормированию трудовых операций посредством использования методов искусственного интеллекта.

*Ключевые слова:* нормирование трудовых операций, непроемственные затраты рабочего времени, компьютерное зрение, искусственный интеллект.

## **COMPUTER VISION IS BASE OF MANUFACTURING WORK TIME LOGGING AUTOMATIZATION**

**M.A. Tsukanov, O.P. Ulyanova**

Stary Oskol technological institute n.a. A.A. Ugarov (branch) National University of Science and  
Technology "MISiS", Stary Oskol  
tsukanov\_m\_a@mail.ru, ulianov032223@mail.ru