# Циклы в структурах функциональных зависимостей

А.А. Карпук, В.В. Краснопрошин

***Аннотация* — Рассматриваются задачи оптимизации структуры функциональных зависимостей (ФЗ) между атрибутами реляционной базы данных: задача построения оптимального элементарного базиса структуры ФЗ между атрибутами и задача поиска оптимального состава таблиц базы данных, находящихся в третьей нормальной форме. Оптимальный элементарный базис структуры ФЗ между атрибутами содержит минимальное количество ФЗ и минимальное количество атрибутов в них. Оптимальный состав таблиц базы данных содержит минимальное количество таблиц. Приведен обзор методов решения перечисленных задач, основанных на поиске $P$ – зависимостей в элементарном базисе структуры ФЗ и поиске порожденных полных подструктур структуры ФЗ, имеющих более одного ключа. Введено понятие цикла в элементарном базисе структуры ФЗ, с использованием которого получены необходимые условия наличия $P$ – зависимостей в элементарном базисе структуры ФЗ и необходимые и достаточные условиях наличия порожденных полных подструктур структуры ФЗ, имеющих более одного ключа. В результате рассматриваемые задачи сведены к задачам поиска и анализа циклов в элементарном базисе структуры ФЗ.**

***Ключевые слова* — реляционная база данных; структура функциональных зависимостей; элементарный базис; ключ; цикл; нормализация таблиц базы данных.**

## I. Введение

Все существующие алгоритмы нормализации таблиц в реляционной базе данных (БД) проводят анализ структуры функциональных зависимостей (ФЗ) между атрибутами, существующих в предметной области. На первом этапе алгоритмов нормализации решается задача удаления тех ФЗ, которые выводятся из остальных ФЗ, а именно задача построения элементарного базиса структуры ФЗ. Эта задача, как и другие задачи анализа структуры ФЗ, формализуется с помощью аксиом Армстронга [1]. При большом количестве атрибутов и ФЗ актуальной становится задача построения элементарного базиса, содержащего минимальное количество ФЗ (минимального элементарного базиса), и задача построения минимального элементарного базиса, содержащего минимальное количество атрибутов (оптимального элементарного базиса). В работе [2] введено понятие $P$ – зависимости в элементарном базисе структуры ФЗ и получены свойства $P$ – зависимостей, позволяющие построить все элементарные базисы структуры ФЗ и выбрать из них минимальные и оптимальные элементарные базисы. Для приведения таблиц реляционной БД к третьей нормальной форме (3НФ) применяются известные алгоритмы Делобеля – Кейси, Бернштейна и Неклюдовой – Цаленко. В работе [3] введено понятие порожденной полной подструктуры структуры ФЗ, с использованием которого получены необходимые и достаточные условия оптимальности результатов работы алгоритмов Делобеля – Кейси и Бернштейна по количеству получаемых таблиц БД. В настоящей работе введено понятие цикла в элементарном базисе структуры ФЗ, с использованием которого получены необходимые условия наличия $P$ – зависимостей в элементарном базисе структуры ФЗ и необходимые и достаточные условиях наличия порожденных полных подструктур структуры ФЗ, имеющих более одного ключа.

Статья получена 1 июня 2017.

А.А. Карпук – кандидат технических наук, доцент, доцент кафедры программного обеспечения сетей телекоммуникаций Белорусской государственной академии связи (e-mail: a_karpuk@mail.ru).

В.В. Краснопрошин – доктор технических наук, профессор, заведующий кафедрой информационных систем управления Белорусского государственного университета (e-mail: krasnoproshin@bsu.by).

## II. Основные определения и обзор предыдущих результатов

**Определение 1** [1]. Пусть $A$ – конечное множество, $X$, $Y$, $C$ и $D$ – подмножества $A$. Будем обозначать $XY = X \bigcup Y$. Структурой ФЗ на множестве $A$ называется бинарное отношение $\rightarrow$ на булеане $B(A)$, удовлетворяющее аксиомам Армстронга:

если $X \subseteq A$, то $A \rightarrow X$ (аксиома рефлексивности);

если $X \rightarrow Y$ и $YC \rightarrow D$, то $XC \rightarrow D$ (аксиома псевдотранзитивности).

Если имеет место ФЗ $X \rightarrow Y$, то говорят, что $X$ функционально определяет $Y$ или $Y$ функционально зависит от $X$. Из приведенных аксиом можно вывести ряд свойств структуры ФЗ, которые в литературе (например, [4]) часто также называются аксиомами, хотя точнее называть их правилами вывода. Наиболее важными являются следующие правила вывода:

если $X \rightarrow YC$, то $X \rightarrow Y$ и $X \rightarrow C$ (правило декомпозиции), действительно, по аксиоме рефлексивности имеем $XYC \rightarrow Y$ и $XYC \rightarrow C$, тогда по аксиоме псевдотранзитивности получим $X \rightarrow Y$ и $X \rightarrow C$;

если $X \rightarrow Y$, то $XC \rightarrow YC$ (правило пополнения), действительно, по аксиоме рефлексивности имеем $XYC \rightarrow YC$, тогда по аксиоме псевдотранзитивности получим $XC \rightarrow YC$;

если $X \rightarrow Y$ и $X \rightarrow C$, то $X \rightarrow YC$ (правило объединения), действительно, по правилу пополнения имеем $X \to XY$ и $XY \to YC$, тогда по аксиоме псевдотранзитивности получим $X \to YC$.

Очевидно, что отношение включения $\subseteq$ определяет структуру ФЗ на $A$, которая называется тривиальной структурой ФЗ, а входящие в нее ФЗ – тривиальными ФЗ. Для задания структуры ФЗ, отличающейся от тривиальной, требуется постулировать конечное множество ФЗ $F = \{F_j = X_j \to Y_j \mid X_j, Y_j \subset A, j = \overline{1,m}\}$, которое в работе [5] было названо системой образующих структуры ФЗ на $A$. Структуру ФЗ, заданную системой образующих $F$, будем обозначать $S(F)$.

**Определение 2** [5]. Замыканием множества $X \subset A$ относительно структуры ФЗ $S(F)$ называется множество $X^+(F) \subseteq A$, такое, что для любого $Y \subseteq A$ из $X \to Y$ следует $Y \subseteq X^+(F)$.

**Определение 3** [6]. Структуры ФЗ $S(F^1)$ и $S(F^2)$ на множестве $A$ с системами образующих $F^1 = \{X_i^1 \to Y_i^1 \mid X_i^1 \subset A, Y_i^1 \subseteq A, i = \overline{1,m_1}\}$ и $F^2 = \{X_j^2 \to Y_j^2 \mid X_j^2 \subset A, Y_j^2 \subseteq A, j = \overline{1,m_2}\}$ соответственно называются эквивалентными, если для любого $X \subset A$ имеет место равенство $X^+(F^1) = X^+(F^2)$.

**Теорема 1** [6]. Структуры ФЗ $S(F^1)$ и $S(F^2)$ на множестве $A$ эквивалентны тогда и только тогда, когда для всех $i = \overline{1,m_1}$ выполняются условия $X_i^{1+}(F^1) = X_i^{1+}(F^2)$ и для всех $j = \overline{1,m_2}$ выполняются условия $X_j^{2+}(F^1) = X_j^{2+}(F^2)$.

**Определение 4** [5]. Система образующих $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(E)$ на множестве $A$ называется элементарным базисом, если выполняются следующие условия:

для любого $j \in \overline{1,m}$ и любого $X \subset T_j$ система образующих $E'$, полученная из $E$ путем замены ФЗ $H_j \to T_j$ на ФЗ $H_j \to T_j \setminus X$, задает структуру ФЗ $S(E')$, не эквивалентную структуре ФЗ $S(E)$;

для любого $j \in \overline{1,m}$ и любого $Y \subset H_j$ система образующих $E'$, полученная из $E$ путем замены ФЗ $H_j \to T_j$ на ФЗ $H_j \setminus Y \to T_j$, задает структуру ФЗ $S(E')$, не эквивалентную структуре ФЗ $S(E)$.

**Определение 5** [6]. Элементарный базис $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(E)$ на множестве $A$ называется минимальным, если не существует другого элементарного базиса структуры ФЗ $S(E)$ с количеством ФЗ, меньшим $m$.

**Определение 6** [6]. Элементарный базис $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(E)$ на множестве $A$ называется оптимальным, если не существует другого минимального элементарного базиса структуры ФЗ $S(E)$ с меньшим количеством вхождений элементов множества $A$.

В англоязычной литературе минимальный и оптимальный элементарный базисы структуры ФЗ называются минимальным и оптимальным покрытием структуры ФЗ [7]. Задача построения минимального и оптимального элементарного базиса структуры ФЗ между атрибутами реляционной БД была поставлена практически одновременно в работе [6] и в англоязычном издании монографии Мейера [7].

**Определение 7** [2]. Пусть $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ – элементарный базис структуры ФЗ $S(E)$ на множестве $A$. В ФЗ $(H_s \to T_s) \in E$ существует $P$ – зависимость непустого $T_s' \subseteq T_s$ от $H_s$, если существует $P \subset H_s^+(E)$, $P \neq H_s$, такое что $T_s' \subseteq [P^+(E) \setminus P]$ и никакое подмножество $P$ этими свойствами не обладает.

Через $E'$ обозначим систему образующих структуры ФЗ, полученную из $E$ путем замены ФЗ $H_s \to T_s$ на $H_s \to (T_s \setminus T_s')$. Различаются три типа $P$ – зависимости: $P_1$ – зависимость имеет место, если $P \subseteq H_s^+(E')$, $P_2$ – зависимость, если одновременно $P \not\subset H_s^+(E')$ и $T_s' \not\subset [P^+(E') \setminus P]$, $P_3$ – зависимость, если одновременно $P \not\subset H_s^+(E')$ и $T_s' \subseteq [P^+(E') \setminus P]$.

**Теорема 2** [2]. Если в ФЗ $(H_s \to T_s) \in E$ существует $P_1$ – зависимость $T_s' \subseteq T_s$ от $H_s$, и система образующих структуры ФЗ $Q$ получена из $E$ путем замены ФЗ $H_s \to T_s$ на $H_s \to (T_s \setminus T_s')$ и добавления к $E$ ФЗ $P \to T_s'$, то структуры ФЗ $S(E)$ и $S(Q)$ эквивалентны.

**Теорема 3** [2]. Если в ФЗ $(H_s \to T_s) \in E$ существует $P_2$ – зависимость $T_s' \subseteq T_s$ от $H_s$, и система образующих структуры ФЗ $Q$ получена из $E$ путем замены ФЗ $H_s \to T_s$ на $H_s \to [(T_s \setminus T_s') \cup (P \setminus H_s)]$ и добавления к $E$ ФЗ $P \to T_s'$, то структуры ФЗ $S(E)$ и $S(Q)$ эквивалентны.

**Теорема 4** [2]. Если в ФЗ $(H_s \to T_s) \in E$ существует $P_3$ – зависимость $T_s' \subseteq T_s$ от $H_s$, и система образующих структуры ФЗ $Q$ получена из $E$ путем замены ФЗ $H_s \to T_s$ на $H_s \to [(T_s \setminus T_s') \cup (P \setminus H_s)]$, то структуры ФЗ $S(E)$ и $S(Q)$ эквивалентны.

Пусть в БД требуется хранить таблицу $R(A_1,..., A_n)$, определенную на множестве атрибутов $A = \{A_1,..., A_n\}$, и задана система образующих $F = \{X_j \to Y_j \mid X_j, Y_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(F)$. Ключом таблицы $R$ называется подмножество атрибутов $K \subseteq A$, такое, что $K^+(F) = A$ и никакое подмножество

$K$ этим свойством не обладает. Атрибут из $A$ называется первичным, если он входит хотя бы в один ключ таблицы $R$. Множество первичных атрибутов из $A$ обозначим $A^p$, множество остальных атрибутов – $A^n$. Таблица $R$ находится в 3НФ, если из $X \to B, X \subset A, B \in A^n, B \notin X$ следует, что $X$ – ключ таблицы $R$. Задача нормализации таблицы $R$ состоит в нахождении проекций $R_1,...,R_q$ таблицы $R$, таких что выполняются следующие условия:

а) $R = R_1 \circ R_2 \circ ... \circ R_q$, где символом $\circ$ обозначена операция естественного соединения проекций;

б) все $R_j, j = \overline{1,q}$, находятся в 3НФ;

в) структура ФЗ $S(F)$ порождается структурами ФЗ проекций $R_j$;

г) проекции $R_j$ зависят только от структуры ФЗ $S(F)$.

Совокупность проекций $R_j$, удовлетворяющих приведенным условиям, называется синтаксическим разложением таблицы $R$. Синтаксическое разложение $R_j, j = \overline{1,q}$ оптимально по количеству проекций, если не существует другого синтаксического разложения $R_1^1,...,R_{q^1}^1$, такого, что $q^1 < q$.

Задача представления таблиц реляционной БД в виде естественного соединения своих проекций, находящихся в 3НФ, была впервые поставлена и решена Коддом. Алгоритм нормализации таблиц БД, предложенный Коддом, имеет существенные недостатки: из получаемых проекций в общем случае не восстанавливается структура ФЗ на множестве атрибутов БД; количество получаемых проекций не является оптимальным. Делобель и Кейси предложили алгоритм, в котором был устранен первый недостаток. Бернштейн разработал алгоритм, обеспечивающий уменьшение количества получаемых проекций. Обобщение алгоритма Бернштейна было выполнено Ислуром. Неклюдова и Цаленко [5] предложили алгоритм, который, по их утверждению, дает оптимальное решение по количеству получаемых проекций. Однако в работе [3] было показано, что в общем случае алгоритм Неклюдовой – Цаленко не дает оптимального решения по количеству получаемых проекций, получены необходимые и достаточные оптимальности алгоритмов Делобеля – Кейси и Бернштейна и предложен алгоритм поиска оптимального решения.

Пусть $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ – оптимальный элементарный базис структуры ФЗ $S(E)$ на множестве $A$. Для произвольного непустого $V \subseteq A$ рассмотрим множество атрибутов $V^+(E)$ как полную подструктуру структуры $S(E)$. Оптимальный элементарный базис этой подструктуры обозначим $E^V$. Полную подструктуру $S(E^V)$ назовем порожденной $V$, множество ее ключевых атрибутов обозначим $V^p$. Выделим в $E^V$ два подмножества ФЗ:

$K = \{H_j \to T_j \in E^V \mid H_j \subset V^p, T_j \subset V^p\}$;

$L = \{H_j \to T_j \in E^V \mid T_j \not\subset V^p$, $H_j$ – ключ подструктуры $S(E^V)\}$.

**Теорема 5** [3]. Алгоритм Делобеля – Кейси дает синтаксическое разложение, оптимальное по количеству проекций, тогда и только тогда, когда для любого непустого $V \subseteq A$ выполняется условие $|K|+|L|<2$. Здесь и далее символом $|M|$ обозначается мощность множества $M$.

**Теорема 6** [3]. Алгоритм Бернштейна дает синтаксическое разложение, оптимальное по количеству проекций, тогда и только тогда, когда для любого непустого $V \subseteq A$, для которого порожденная полная подструктура $S(E^V)$ имеет более одного ключа и выполняется условие $|K|+|L|<2$, все ключи порожденной полной подструктуры $S(E^V)$ принадлежат множеству левых частей элементарного базиса $E$.

Алгоритм синтеза таблиц БД в 3НФ, дающий оптимальное решение по количеству проекций, состоит из следующих шагов [3].

1. Построить оптимальный элементарный базис $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(E)$.

2. Найти непустое $V \subseteq A$, такое, что порожденная полная подструктура $S(E^V)$ структуры ФЗ $S(E)$, имеет более одного ключа, и выполняется условие $|K|+|L| \geq 2$. Если такого подмножества атрибутов нет, то перейти к шагу 4. Если таких подмножеств атрибутов более одного, то выбрать подмножество с максимальным значением величины $|K|+|L|$.

3. В формируемое решение включить проекцию таблицы $R$ на множество атрибутов, полученное добавлением к $V^p$ правых частей зависимостей из $L$. Удалить из $E$ все ФЗ, вошедшие в построенную проекцию. Если в $E$ осталось более одной ФЗ, то перейти к шагу 2.

4. Для каждой ФЗ, оставшейся в $E$, в формируемое решение включить проекцию таблицы $R$ на множество атрибутов этой ФЗ. Если хотя бы одна из проекций, полученных на шагах 3 и 4, содержит один из ключей таблицы $R$, то закончить работу, иначе в формируемое решение добавить проекцию таблицы $R$ на множество атрибутов одного из ключей.

Задачу построения оптимального элементарного базиса структуры ФЗ на шаге 1 алгоритма можно решить путем построения элементарного базиса с использованием алгоритма, описанного в работе [2], поиска всех $P$ – зависимостей в элементарном базисе и его оптимизации с использованием теорем 2 – 4. На шаге 2 алгоритма требуется решить задачу поиска всех порожденных полных подструктур структуры ФЗ, имеющих более одного ключа.

## III. Циклы в структурах функциональных зависимостей и их свойства

**Определение 8**. Циклом в элементарном базисе $E=\{H_j \to T_j \mid H_j, T_j \subset A, j=\overline{1,m}\}$ структуры ФЗ $S(E)$ на множестве $A$ назовем последовательность ФЗ $\{H_{j_1} \to T_{j_1},...,H_{j_k} \to T_{j_k} \mid H_{j_\nu} \to T_{j_\nu} \in E, \nu=\overline{1,k}\}$, таких что $S_{j_1}=H_{j_1} \bigcap T_{j_k} \neq \varnothing$, $S_{j_\nu}=H_{j_\nu} \bigcap T_{j_{\nu-1}} \neq \varnothing, \nu=\overline{2,k}$, причем для любых $\alpha,\beta \in \overline{1,k}, \alpha \neq \beta$, выполняется условие $S_{j_\alpha} \bigcap S_{j_\beta}=\varnothing$.

**Теорема 7.** Пусть $E=\{H_j \to T_j \mid H_j, T_j \subset A, j=\overline{1,m}\}$ – элементарный базис структуры ФЗ $S(E)$ на множестве $A$. Если в ФЗ $H_s \to T_s \in E$ существует $P$ – зависимость непустого $T_s' \subseteq T_s$ от $H_s$, то в $E$ существует цикл.

Доказательство. Пусть в ФЗ $H_s \to T_s$ существует $P$ – зависимость $T_s' \subseteq T_s$ от $H_s$. Обозначим $H_s'=H_s \setminus P$, $P'=P \setminus H_s$. Из $P \subset H_s^+(E)$ следует, что для любого $A' \in P'$ существует последовательность ФЗ $H_{j_1} \to T_{j_1},...,H_{j_p} \to T_{j_p}$, принадлежащих $E$, таких что $H_{j_1}=H_s$, $T_{j_1}=T_s$, $A' \in T_{j_p}$, $H_{j_\nu} \bigcap T_{j_{\nu-1}} \neq \varnothing, \nu=\overline{2,p}$.

Из $H_s \subset P^+(E')$ для $P_1$ – зависимости и $H_s \subset P^+(E)$ для $P_2$ – зависимости следует, что для любого $B' \in H_s'$ существует последовательность ФЗ $H_{l_1} \to T_{l_1},...,H_{l_q} \to T_{l_q}$ принадлежащих $E$, таких что $H_{l_1} \subseteq P$, $B' \in T_{l_q}$, $H_{l_\nu} \bigcap T_{l_{\nu-1}} \neq \varnothing, \nu=\overline{2,q}$. Очевидно, что $H_{l_1} \bigcap P' \neq \varnothing$, и построенные последовательности ФЗ образуют цикл, в состав которого входит ФЗ $H_s \to T_s$.

Пусть имеет место $P_3$ – зависимость $T_s' \subseteq T_s$ от $H_s$. Для определенной выше последовательности ФЗ $H_{j_1} \to T_{j_1},...,H_{j_p} \to T_{j_p}$ существует минимальный индекс $\gamma$, для которого выполняется условие $H_{j_\gamma} \bigcap T_s' \neq \varnothing$. Из $T_s' \subseteq [P^+(E') \setminus P]$ следует, что для любого $C' \in T_s'$ существует последовательность ФЗ $H_{l_1} \to T_{l_1},...,H_{l_q} \to T_{l_q}$ принадлежащих $E$, таких что $H_{l_1} \subseteq P$, $C' \in T_{l_q}$, $H_{l_\nu} \bigcap T_{l_{\nu-1}} \neq \varnothing, \nu=\overline{2,q}$. Таким образом, последовательность ФЗ $H_{j_\gamma} \to T_{j_\gamma},...,H_{j_p} \to T_{j_p}, H_{l_1} \to T_{l_1},...,H_{l_q} \to T_{l_q}$ образует цикл, в который не входит ФЗ $H_s \to T_s$. Теорема доказана.

**Теорема 8.** Пусть $E=\{H_j \to T_j \mid H_j, T_j \subset A, j=\overline{1,m}\}$ – элементарный базис структуры ФЗ $S(E)$ на множестве $A$. Если для некоторого непустого $V \subseteq A$ порожденная полная подструктура $S(E^V)$ структуры ФЗ $S(E)$ имеет более одного ключа, то в $E^V$ существует цикл.

Доказательство. Пусть $K_1 \subset V$ и $K_2 \subset V$ – ключи порожденной полной подструктуры $S(E^V)$. Обозначим $K_1'=K_1 \setminus K_2$, $K_2'=K_2 \setminus K_1$. Из $K_2' \subseteq K_1^+(E^V)$ следует, что для любого $A' \in K_2'$ существует последовательность ФЗ $H_{j_1} \to T_{j_1},...,H_{j_p} \to T_{j_p}$, принадлежащих $E^V$, таких что $H_{j_1} \subseteq K_1$, $A' \in T_{j_p}$, $H_{j_\nu} \bigcap T_{j_{\nu-1}} \neq \varnothing, \nu=\overline{2,p}$. Существуют $A' \in K_2'$ и минимальный индекс $\gamma \in \{1,2,...,p\}$, для которых выполняется условие $H_{j_\gamma} \bigcap K_1' \neq \varnothing$, поскольку в противном случае ключом порожденной полной подструктуры $S(E^V)$ является $K_1 \setminus K_1'$. Из $K_1' \subseteq K_2^+(E^V)$ следует, что для любого $B' \in H_{j_\gamma} \bigcap K'$ существует последовательность ФЗ $H_{j_{p+1}} \to T_{j_{p+1}},...,H_{j_k} \to T_{j_k}$, принадлежащих $E^V$, таких что $H_{j_{p+1}} \subseteq K_2$, $B' \in T_{j_k}$, $H_{j_\nu} \bigcap T_{j_{\nu-1}} \neq \varnothing, \nu=\overline{p+2,k}$. Если для всех $B' \in H_{j_\gamma} \bigcap K'$ имеет место $A' \notin H_{j_p}, \nu=\overline{p+1,k}$, то структуре ФЗ $S(E^V)$ принадлежит ФЗ $K_2 \setminus A' \to A'$ и $K_2$ не является ключом, что противоречит исходному предположению. Следовательно, для некоторого $B' \in H_{j_\gamma} \bigcap K'$ существует $\varepsilon \in \overline{p+1,k}$, такое что $A' \in H_{j_\varepsilon}$ и последовательность ФЗ $H_{j_\gamma} \to T_{j_\gamma},...,H_{j_p} \to T_{j_p}, H_{j_\varepsilon} \to T_{j_\varepsilon},...,H_{j_k} \to T_{j_k}$ образует цикл. Теорема доказана.

Таким образом, наличие цикла в элементарном базисе структуры ФЗ является необходимым условием существования $P$ – зависимостей в элементарном базисе структуры ФЗ и существования порожденных полных подструктур структуры ФЗ, имеющих более одного ключа. Однако данное условие не является достаточным, что показывает следующий пример.

**Пример 1**. Пусть $A=\{A_1,A_2,A_3,A_4,A_5\}$, таблица $R$ определена на всех атрибутах из $A$ и $F=\{A_1A_2 \to A_3, A_1A_3 \to A_4, A_2A_4 \to A_5, A_5 \to A_3\}$. Элементарный базис $E=F$, единственным ключом является $K_1=A_1A_2$. Последние три ФЗ элементарного базиса образуют цикл, однако в элементарном базисе отсутствуют $P$ – зависимости, и для любого непустого $V \subseteq A$ порожденная полная подструктура $S(E^V)$ структуры ФЗ $S(E)$ имеет один ключ.

Для упрощения обозначений будем считать, что в элементарном базисе $E=\{H_j \to T_j \mid H_j, T_j \subset A, j=\overline{1,m}\}$ структуры $S(E)$ на множестве $A$ цикл образует последовательность ФЗ $\{H_1 \to T_1,...,H_k \to T_k \mid H_\nu \to T_\nu \in E, \nu = \overline{1,k}\}$, что всегда можно достичь путем перенумерации ФЗ.

**Определение 9**. Окрестностью цикла $\{H_1 \to T_1,...,H_k \to T_k \mid H_\nu \to T_\nu \in E, \nu = \overline{1,k}\}$ в элементарном базисе $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры $S(E)$ на множестве $A$ назовем множество $D = \bigcup_{\nu=1}^{k}(H_\nu \setminus S_\nu)$.

Для цикла $\{H_1 \to T_1,...,H_k \to T_k \mid H_\nu \to T_\nu \in E, \nu = \overline{1,k}\}$ построим множество $V = \bigcup_{\nu=1}^{k} H_\nu$ и рассмотрим множество атрибутов $V^+(E)$ как полную подструктуру структуры $S(E)$. Элементарный базис этой подструктуры обозначим $E^V$, очевидно, что $E^V \subseteq E$, причем в состав $E^V$ входят все ФЗ из рассматриваемого цикла.

**Лемма 1.** Если для некоторого $\nu \in \overline{1,k}$ ФЗ $D \to S_\nu$ принадлежат порожденной полной подструктуре $S(E^V)$, то для всех $\nu = \overline{1,k}$ ФЗ $D \to S_\nu$ принадлежат порожденной полной подструктуре $S(E^V)$ и эта подструктура имеет единственный ключ.

Утверждение леммы легко доказывается с использованием определения цикла в элементарном базисе структуры ФЗ. Условиям леммы удовлетворяет цикл из примера 1, поэтому порожденная полная подструктура $S(E^V)$ в примере 1 имеет единственный ключ.

**Теорема 9.** Если для некоторого $\nu \in \overline{1,k}$ ФЗ $D \to S_\nu$ не принадлежит порожденной полной подструктуре $S(E^V)$, то порожденная полная подструктура $S(E^V)$ имеет $k$ ключей $K_\nu = S_\nu \cup D$, $\nu = \overline{1,k}$.

Доказательство. Если выполняется условие теоремы, то из леммы 1 следует, что для всех $\nu = \overline{1,k}$ ФЗ $D \to S_\nu$ не принадлежат порожденной полной подструктуре $S(E^V)$. В силу наличия цикла $\{H_1 \to T_1,...,H_k \to T_k \mid H_\nu \to T_\nu \in E, \nu = \overline{1,k}\}$ и определения окрестности $D$ для любого $\nu \in \overline{1,k}$ имеем $V \subseteq K_\nu^+(E^V)$, т. е. $K_\nu$ – ключ порожденной полной подструктуры $S(E^V)$, из которого нельзя удалить элементы множества $S_\nu$. Поскольку по определению цикла все множества $S_\nu$ попарно различны, то и все ключи $K_\nu$, $\nu \in \overline{1,k}$, попарно различны.

## IV. Заключение

Таким образом, задачи поиска $P$ – зависимостей в элементарном базисе структуры ФЗ и поиска всех порожденных полных подструктур структуры ФЗ, имеющих более одного ключа, сводятся к поиску и анализу циклов в элементарном базисе структуры ФЗ. Элементарному базису $E = \{H_j \to T_j \mid H_j, T_j \subset A, j = \overline{1,m}\}$ структуры ФЗ $S(E)$ на множестве $A$ можно поставить в соответствие двудольный ориентированный граф $(A, E, H, T)$, в котором $A$ – множество вершин первой доли графа, $E$ – множество вершин второй доли графа, $H$ – множество дуг графа, направленных от вершин первой доли к вершинам второй доли графа (показывают вхождение элементов из $A$ в левые части ФЗ из $E$), $H$ – множество дуг графа, направленных от вершин второй доли к вершинам первой доли графа (показывают вхождение элементов из $A$ в правые части ФЗ из $E$). Легко убедиться, что каждому циклу в элементарном базисе $E$ соответствует семейство циклов в соответствующем двудольном графе. Количество циклов в семействе равно произведению мощностей множеств $S_\nu$, $\nu = \overline{1,k}$. В частности, если все множества $S_\nu$ содержат по одному элементу, то такому циклу в элементарном базисе $E$ соответствует единственный цикл в соответствующем двудольном графе. В общем случае задача поиска всех циклов в двудольном ориентированном графе является $NP$ – трудной, но ее $NP$ – трудность определяется тем, что максимальное возможное количество циклов в графе экспоненциально зависит от размерности графа. Время поиска одного цикла в ориентированном графе стандартным алгоритмом поиска в глубину линейно зависит от размерности графа. В реальных задачах проектирования реляционных БД при количестве атрибутов порядка $10^3$ количество циклов в элементарном базисе структуры ФЗ не превосходит $10^2$, поэтому все циклы в элементарном базисе структуры ФЗ можно найти за приемлемое время.

## Библиография


[1] W.W. Armstrong. Dependency structure of data base relationships // Proc. IFIP Congress. Geneva, Switzerland, 1974. P. 580–583.

[2] А.А. Карпук. О построении элементарного базиса системы функциональных зависимостей в базе данных // Информационные технологии и программные средства: проектирование, разработка и применение: сб. научн. статей / Под ред. Л.В. Рудиковой. Гродно: ГрГУ, 2011. С. 185–190.

[3] А.А. Карпук. Алгоритмы нормализации таблиц реляционной базы данных // Системы управления и информационные технологии. 2017. № 2(68). С. 53–57.

[4] С.Д. Кузнецов. Базы данных. Модели и языки. М: Бином-Пресс, 2008. 720 с.

[5] Е.А. Неклюдова, М.Ш. Цаленко. Синтез логической схемы реляционной базы данных // Программирование. 1979. № 6. С. 58–68.

[6] А.А. Карпук. Выбор элементарного базиса структуры функциональных зависимостей при проектировании базы данных // Вопросы радиоэлектроники. Сер. ОВР. 1983. Вып. 6. С. 38–41.

[7] Д. Мейер. Теория реляционных баз данных: Пер. с англ. М: Мир, 1987. 608 с.



**Анатолий Алексеевич Карпук** окончил факультет прикладной математики Белорусского государственного университета в 1974 г. В 1985 г. защитил в Центральном НИИ Министерства обороны СССР (г. Москва) диссертацию на соискание ученой степени кандидата технических наук. В 1991 г. решением ВАК СССР ему присвоено ученое звание старшего научного сотрудника, а в 2008 г. решением ВАК Республики Беларусь – ученое звание доцента. Ведет научные

исследования в областях моделирования и оптимизации сложных систем, проектирования баз данных, оценки качества радиосвязи и оптимизации присвоения частот радиолиниям. По результатам исследований опубликовал 2 монографии и около 120 научных трудов.

С 1974 по 2003 г. он работал в НПО "Агат" в г. Минске, где прошел путь от инженера – стажера до главного конструктора программных средств специального назначения. Под его руководством была разработана одна из первых в мировой практике объектно – реляционная СУБД специального назначения и логическая и физическая структуры баз данных АСУ войсками тактического и оперативного звеньев управления "Маневр". В 2003 –2016 г. он работал начальником управления в Расчетном центре Национального банка Республики Беларусь, где руководил разработкой и оптимизацией баз данных ряда подсистем автоматизированной системы межбанковских расчетов. С 2016 г. он работает доцентом кафедры программного обеспечения сетей телекоммуникаций в Белорусской государственной академии связи и продолжает работать в ОАО "Агат – системы управления" (ранее – НПО "Агат") в должности ведущего научного сотрудника.

В 1992 г. Карпук А.А. был избран гранд – инженером (членом – корреспондентом) Белорусской инженерно – технологической академии.

**Виктор Владимирович Краснопрошин** окончил факультет прикладной математики Белорусского государственного университета (БГУ) в 1974 г. и аспирантуру БГУ в 1979 г. В 1979 г. защитил в БГУ диссертацию на соискание ученой степени кандидата физико – математических наук и в 2007 г. – диссертацию на соискание ученой степени доктора технических наук. В 1983 г. решением ВАК СССР ему присвоено ученое звание доцента, а в 2008 г. решением ВАК Республики Беларусь – ученое звание профессора. Ведет научные исследования в областях распознавания образов, искусственного интеллекта, теории принятия решений, информационно – компьютерных технологий. По результатам исследований опубликовал 7 монографий и около 270 научных трудов.

С 1974 по настоящее время он работает в БГУ в г. Минске, где прошел путь от стажера – преподавателя до заведующего кафедрой информационных систем управления. Руководит научно – исследовательской лабораторией "Информационные технологии и компьютерная графика", которая на протяжении ряда лет выполняет проекты по важнейшим республиканским и союзным научно – техническим программам: "Научные основы информационных технологий и систем", "Защита от чрезвычайных ситуаций", "Информатика и космос", "Недра" и другим. Разработанные системы внедрены в различных организациях и предприятиях Республики Беларусь.

В настоящее время Краснопрошин В.В. является иностранным академиком Испанской Королевской Академии экономики и финансов; вице – президентом Белорусской ассоциации по распознаванию и анализу изображений; членом управляющих советов международных научных ассоциаций AEDEM, AMSE и SIGEF; членом государственного экспертного совета Республики Беларусь по информатизации, вычислительной технике и информационным технологиям; членом экспертного совета фонда фундаментальных исследований НАН Республики Беларусь по информатике; членом экспертного совета ВАК Республики Беларусь по техническим наукам; членом специализированного совета по защите докторских диссертаций.

# Cycles in Structures of functional Dependencies

A.A. Karpuk, V.V. Krasnoproshin

***Abstract* — The main tasks of optimization of structure of the functional dependences (FD) between attributes of a relational database are considered: the task of creation of an optimum elementary base of structure of the FD between attributes and the task of search of optimum composition of the tables of the database which are in the third normal form. The optimum elementary base of structure of the FD between attributes contains the minimum quantity of the FD and the minimum quantity of attributes in them. The optimum composition of tables of the database contains the minimum quantity of tables. The review of methods of the decision of tasks of optimization of structure of the FD between attributes based on search of $P$– dependences in an elementary base of structure of the FD and search of the generated complete substructures of structure of the FD having more than one key is provided. The concept of a cycle of an elementary base of structure of the FD is entered. Necessary conditions of existence of $P$– dependences in an elementary base of structure of the FD both necessary and sufficient conditions of existence of the generated complete substructures of structure of the FD having more than one key are received. As a result the considered tasks are consolidated to tasks of search and the analysis of cycles in an elementary base of structure of the FD.**

***Keywords* — relational database; structure of functional dependences; elementary base; key; cycle; normalization of database tables.**