

Вісник Харківського національного університету
Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи
управління»
УДК 004.652/.942 № 629, 2004, с.11-24

Представление полусхем предметных областей информационных систем средствами реляционных баз данных

Г. Н. Жолткевич, Т. В. Семенова, К. А. Федорченко
Харьковский национальный университет им. В.Н. Каразина, Украина
Херсонский государственный университет, Украина

Method of presentation of semi-schemes from [1] using relational databases was suggested in given article. Algorithms of correctness verification of a data domain conceptual model corresponding to the semi-scheme were suggested and proved. It gives opportunity of the program realization of the given method in the form of CASE-tools for conceptual modeling.

Традиционная парадигма построения информационных систем базируется на ряде предположений, которые исключают возможность адаптации системы к структурным изменениям знаний о предметной области. Статичность концептуальной модели предметной области является одним из таких предположений. Преодоление трудностей, связанных с этим предположением, требует введение в информационную систему уровня метаданных, представляющих семантику предметной области [2]. Представление метаданных, по возможности, должно быть реализовано с использованием тех же механизмов хранения и доступа, которые используются в системе для обработки данных.

В работе [1] авторами предложен алгебраический подход к построению концептуальных моделей предметных областей, базирующийся на введенном ими классе полусхем. Этот подход позволил построить алгоритмы проверки корректности концептуального описания предметной области с точки зрения существования образцов понятий, определяемых полусхемой. Полусхема, корректная в том смысле, что все определяемые ее понятия имеют не менее одного образца, названа авторами схемой предметной области. Таким образом, схема может рассматриваться как концептуальная модель предметной области, поскольку фиксирует только структурные отношения между понятиями.

Настоящая работа посвящена дальнейшему исследованию полусхем и построению их представлений в терминах теории реляционных баз данных. Выбор указанного представления обусловлен, в первую очередь, технологической зрелостью реляционного подхода, что позволяет при незначительных затратах строить реализации алгоритмов, обеспечивающих проверку корректности моделей, а также интегрировать концептуальный уровень описания данных в базу данных информационной системы.

Мы надеемся, что предлагаемый подход позволит создать технологию, которая предоставит не только возможность разрабатывать и внедрять в

эксплуатацию информационную систему в кратчайшие сроки, но и обеспечит механизмы гибкой настройки системы с целью сохранения ее актуальности в течение как можно большего промежутка времени.

Также как и в работе [1] будем использовать следующие обозначения.

N – конечное множество, элементы которого соответствуют именам понятий предметной области.

R – конечное множество, элементы которого соответствуют именам ролей, т.е. ссылкам внутри экземпляра понятия на его структурные части.

Для пары множеств X, Y обозначим

$M_+(X, Y)$ – множество частичных, хотя бы где-то определенных, отображений из X в Y ;

$dom(f)$ – область определения отображения $f \in M_+(X, Y)$;

$im(f)$ – область значения отображения $f \in M_+(X, Y)$.

Определение 1.

Полусхемой предметной области назовем тройку $S = (N, R, D)$, где N, R – конечные множества, $D \subset N \times M_+(R, N)$, для которой выполняется условие:

для $n \in N, f, g \in M_+(R, N), r \in R$ таких, что $(n, f) \in D, (n, g) \in D$ и $r \in dom(f) \cap dom(g)$, верно $f(r) = g(r)$.

Приведенное определение 1 несколько отличается от соответствующего определения в работе [1], однако несложно увидеть, что все результаты этой работы сохраняются при такой модификации определения полусхемы.

В связи с тем, что было модифицировано определение полусхемы, определение базового понятия также требует модификации.

Определение 2.

Для полусхемы $S = (N, R, D)$ понятие $n \in N$ называется **базовым**, если для всякого $f \in M_+(R, N)$ выполняется $(n, f) \notin D$.

Определение 3.

Пусть тройка $S = (N, R, D)$ является полусхемой и для некоторого $n \in N$ существует $f \in M_+(R, N)$ такая, что $(n, f) \in D$. Тогда будем говорить, что для понятия n задан **вариант определения** f .

Определение 4.

Приведенной полусхемой предметной области назовем полусхему $S = (N, R, D)$, для которой:

1. $\forall (n \in N) \exists (f \in M_+(R, N)) ((n, f) \in D) \vee \exists (r \in dom f) f(r) = n$;
2. $\forall (r \in R) \exists (n \in N) \exists (f \in M_+(R, N)) ((n, f) \in D) r \in dom f$.

В дальнейшем, не теряя общности, можно считать, что все полусхемы являются приведенными. В противном случае не трудно указать эффективный метод, позволяющий перейти от не приведенной полусхемы к приведенной.

Зададим инъективное отображение $name : M_+(R, N) \rightarrow \Sigma$, где Σ – некоторое множество, элементы которого используются для именования частично-определенных функций.

Свяжем с каждой полусхемой $S = (N, R, D)$ отношение m_S со схемой [3] $M(T : N, C : \Sigma, P : R, V : N)$, определяемое следующим образом:

$$\langle n, c, r, n' \rangle \in m_S \Leftrightarrow \exists (f \in M_+(R, N) \mid name(f) = c) \\ ((n, f) \in D \wedge r \in dom(f) \wedge f(r) = n')$$

Лемма 1.

Для отношения $m_S : M(T : N, C : \Sigma, P : R, V : N)$, построенного по полусхеме $S = (N, R, D)$, выполняются следующие зависимости:

1. $C \twoheadrightarrow T$ (C многозначно определяет T [3]);
2. $CP \rightarrow V$ (C и P функционально определяют V [3]);
3. $TP \rightarrow V$.

Доказательство.

Проверим выполнение зависимости 1.

Пусть $t_1 = \langle n_1, c, r_1, n'_1 \rangle \in m_S$ и $t_2 = \langle n_2, c, r_2, n'_2 \rangle \in m_S$. В силу определения m_S имеем:

Существует $f \in M_+(R, N)$ такая, что $name(f) = c$ и

$$(n_1, f) \in D \wedge r_1 \in dom(f) \wedge f(r_1) = n'_1$$

$$(n_2, f) \in D \wedge r_2 \in dom(f) \wedge f(r_2) = n'_2$$

Рассмотрим кортеж $t_3 = \langle n_1, c, r_2, n'_2 \rangle$ и покажем, что он также принадлежит m_S .

Действительно, из условия 1) получаем $(n_1, f) \in D$, а из условия 2) следует, что $r_2 \in dom(f) \wedge f(r_2) = n'_2$, что и означает $t_3 \in m_S$. Таким образом, зависимость 1 выполнена.

Проверим выполнение зависимости 2.

Пусть $t_1 = \langle n_1, c, r, n'_1 \rangle \in m_S$ и $t_2 = \langle n_2, c, r, n'_2 \rangle \in m_S$, тогда в силу определения отношения m_S существует $f \in M(R, N)$ с именем c , причем $f(r) = n'_1$ и $f(r) = n'_2$, а значит $n'_1 = n'_2$. Таким образом, зависимость 2 выполнена.

Проверим выполнение зависимости 3.

Пусть $t_1 = \langle n, c_1, r, n'_1 \rangle \in m_S$ и $t_2 = \langle n, c_2, r, n'_2 \rangle \in m_S$, тогда в силу определения отношения m_S и характеристического свойства полусхемы получаем $n'_1 = n'_2$. Таким образом, зависимость 3 выполнена, что и завершает доказательство леммы.

Лемма 2.

Для отношения $m_S : M(T : N, C : \Sigma, P : R, V : N)$, построенного по полусхеме $S = (N, R, D)$, выполняется свойство:

$$\left(\forall c', c'' \in \text{im}(\text{name}) \mid \pi_{PV} \left(\sigma_{C=c'}(m_S) \right) = \pi_{PV} \left(\sigma_{C=c''}(m_S) \right) \right) \quad c' = c''.$$

Доказательство.

Введем обозначение $\Gamma_c = \pi_{PV} \left(\sigma_{C=c}(m_S) \right)$. В силу построения отношения m_S существует такая функция $f \in M_+(R, N)$, что $\text{name}(f) = c$. Существование такой функции следует из условия, что $c \in \text{im}(\text{name})$, а единственность является следствием инъективности отображения name .

Покажем, что множество Γ_c является графиком найденной функции f . Пусть $\Gamma_f = \{(r, n) \mid r \in \text{dom}(f) \wedge f(r) = n\}$ – график функции f . Убедимся, что $\Gamma_c = \Gamma_f$. Докажем включение $\Gamma_c \subset \Gamma_f$.

Пусть $t = \langle r, n \rangle \in \Gamma_c$, тогда в силу построения отношения m_S выполнено $\langle r \in \text{dom}(f) \wedge f(r) = n \rangle$, а значит, $\langle r, n \rangle \in \Gamma_f$. Следовательно, $\Gamma_c \subset \Gamma_f$.

Обратно. Пусть пара $\langle r, n' \rangle \in \Gamma_f$. Тогда, поскольку полусхема $S = (N, R, D)$ является приведенной, существует понятие $n \in N$ такое, что кортеж $t = \langle n, c, r, n' \rangle \in m_S$. А следовательно, $t' = \langle r, n' \rangle \in \Gamma_c$. Включение $\Gamma_c \supset \Gamma_f$ доказано. Следовательно, $\Gamma_c = \Gamma_f$.

Доказательство завершено.

Лемма 3.

Пусть заданы приведенные полусхемы $S_1 = (N_1, R_1, D_1)$, $S_2 = (N_2, R_2, D_2)$ и по ним построены отношения m_{S_1} и m_{S_2} . Тогда из $m_{S_1} = m_{S_2}$ следует совпадение полусхем S_1 и S_2 .

Доказательство.

Проведем доказательство от противного. Пусть $m_{S_1} = m_{S_2}$, но $S_1 \neq S_2$.

Полусхемы S_1 и S_2 не совпадают, если выполнено хотя бы одно из трех утверждений: либо $N_1 \neq N_2$, либо $R_1 \neq R_2$, либо $D_1 \neq D_2$. Рассмотрим каждый из трех случаев отдельно.

Предположим, что $N_1 \neq N_2$. Покажем, что такое предположение влечет противоречие условию $m_{S_1} = m_{S_2}$. Для определенности будем считать, что существует понятие $n \in N_1$ такое, что $n \notin N_2$. Возможны два случая:

1. Понятие n – базовое. Тогда в силу приведенности S_1 существует некоторое понятие $n' \in N_1$ для которого найдется такая функция $f \in M_+(R, N)$, что

$(n', f) \in D$ и $\exists(r \in \text{dom } f)(f(r) = n)$. Тогда, в силу построения, отношение m_{S_1} содержит кортеж $t = \langle n', c, r, n \rangle$, где $c = \text{name}(f)$.

2. Понятие n не является базовым. Тогда в силу приведенности S_1 $\exists(f \in M_+(R, N))((n, f) \in D_1 \wedge (\exists r \in \text{dom } f) f(r) = n')$. Следовательно, отношение m_{S_1} содержит кортеж $t = \langle n, c, r, n' \rangle$, где $c = \text{name}(f)$.

Очевидно, что и в первом, и во втором случае кортеж t не может принадлежать отношению m_{S_2} . Следовательно, $m_{S_1} \neq m_{S_2}$, что противоречит условию леммы.

Аналогичные рассуждения приводят к противоречию в случае $R_1 \neq R_2$.

Рассмотрим случай, когда $N_1 = N_2$, $R_1 = R_2$, но $D_1 \neq D_2$. Покажем, что такое предположение также противоречит условию $m_{S_1} = m_{S_2}$. Для определенности будем считать, что существует такое понятие $n \in N_1 = N_2$, что $(n, f) \in D_1$, но $(n, f) \notin D_2$. Пусть $r \in \text{dom } f \wedge f(r) = n'$. Тогда кортеж $t = \langle n, c, r, n' \rangle$, где $c = \text{name}(f)$, принадлежит отношению m_{S_1} , но не m_{S_2} . Следовательно, $m_{S_1} \neq m_{S_2}$, что противоречит условию леммы.

Рассмотрены все три случая и в каждом из них получено противоречие с исходным предположением, что и доказывает утверждение.

Теорема 1.

Для всякого отношения m со схемой $M(T : N, C : \Sigma, P : R, V : N)$, удовлетворяющего зависимостям

1. $C \rightarrow T(C \rightarrow PV)$;
2. $CP \rightarrow V$;
3. $TP \rightarrow V$,

и для которого выполнено свойство

$$4. \forall (c', c'' \in \text{im}(\text{name}) \mid \pi_{PV}(\sigma_{C=c'}(m_S)) = \pi_{PV}(\sigma_{C=c''}(m_S))) c' = c''$$

существует единственная приведенная полусхема $S = (N, R, D)$ такая, что $m_S = m$.

Доказательство.

Построим полусхему $S = (N, R, D)$ по заданному отношению m со схемой $M(T : N, C : \Sigma, P : R, V : N)$, удовлетворяющему приведенным зависимостям.

В качестве множества N полусхемы S возьмем множество значений, принимаемых атрибутами T и V отношения $m : N = \pi_T(m) \cup \pi_V(m)$.

В качестве множества R полусхемы S возьмем множество значений, принимаемых атрибутом P отношения $m : R = \pi_P(m)$.

На основе полученных N и R построим множество частичных, хотя бы где-то определенных отображений $M_+(R, N)$.

Зададим отображение $name : M_+(R, N) \rightarrow \Sigma$, где Σ - множество, элементы которого будут именовать частично-определенные функции. При этом, предполагается, что $\pi_C(m) \subset \Sigma$. Будем говорить, что $name(f) = c$, если $\Gamma_c = \pi_{PV}(\sigma_{C=c}(m))$ является графиком функции f .

Однозначность отображения $name : M_+(R, N) \rightarrow \Sigma$ следует из условия 4 теоремы.

Тогда определим множество D следующим образом:

$$D = \left\{ (n, f) \in N \times M_+(R, N) \mid \exists (c \in \pi_C(m)) (name(f) = c \wedge t = \langle n, c \rangle \in \pi_{TC}(m)) \right\}$$

В силу многозначной зависимости $C \rightarrow\rightarrow T(C \rightarrow\rightarrow PV)$ имеем:

если в отношении m присутствуют оба кортежа $\langle n_1, f, r_1, n' \rangle$, $\langle n_2, f, r_2, n'' \rangle$, то присутствуют также оба кортежа $\langle n_2, f, r_1, n' \rangle$, $\langle n_1, f, r_2, n'' \rangle$. Что обеспечивает корректное задание отношения $D \subset N \times M_+(R, N)$.

Осталось проверить, что для построенной тройки $S = (N, R, D)$ выполнено свойство полусхемы.

В силу функциональной зависимости $TP \rightarrow V$ имеем:

$$\forall (t_1, t_2 \in m) (t_1 = \langle n, c_1, r, n' \rangle, t_2 = \langle n, c_2, r, n'' \rangle \Rightarrow n' = n'').$$

Следовательно, независимо от значения второго элемента кортежа понятие в каждой роли определено однозначно. Что и обеспечивает выполнение свойства полусхемы.

Докажем, что $m_S = m$, где m_S - отношение, построенное на основе полусхемы $S = (N, R, D)$.

Пусть $t = \langle n, c, r, n' \rangle \in m$. Тогда, в силу построения, для полусхемы S , получаемой на основе отношения m , будет выполнено следующее: $n \in N$, $(n, f) \in D$, $f(r) = n'$, $r \in \text{dom}(f)$, где $name(f) = c$. Но выписанные условия, в свою очередь, являются необходимыми и достаточными, чтобы кортеж $t = \langle n, c, r, n' \rangle$ принадлежал отношению m_S . Таким образом, выполнено включение $m \subset m_S$.

Докажем обратное включение.

Пусть $t = \langle n, c, r, n' \rangle \in m_S$. Это выполнено тогда и только тогда, когда для полусхемы S выполнено следующее: $n \in N$, $(n, f) \in D$, $r \in \text{dom}(f)$, $f(r) = n'$, где $name(f) = c$. Поскольку, полусхема S была построена на основе отношения m , очевидно, что кортеж $t = \langle n, c, r, n' \rangle \in m$. Таким образом, выполнено включение $m \supset m_S$, а значит $m_S = m$.

Единственность полусхемы $S = (N, R, D)$, построенной по отношению m , немедленно следует из леммы 3.

Теорема 2.

Если отношение m со схемой $M(T : N, C : \Sigma, P : R, V : N)$ удовлетворяет множеству зависимостей $\{C \rightarrow\rightarrow T, CP \rightarrow V, TP \rightarrow V\}$, тогда имеет место

соединение без потерь $m = \pi_{TC}(m) \triangleright \triangleleft \pi_{CPV}(m) \triangleright \triangleleft \pi_{TPV}(m)$, причем ключами для проекций отношения m являются соответственно TC , CP и TP .

Доказательство.

В силу теоремы Фейгина [3] наличие многозначной зависимости $C \rightarrow \rightarrow T(C \rightarrow \rightarrow PV)$ позволяет представить отношение m в виде естественного соединения двух проекций: $m = \pi_{TC}(m) \triangleright \triangleleft \pi_{CPV}(m)$. Однако, в этом случае возможна потеря функциональной зависимости $TP \rightarrow V$. Поскольку естественное соединение отношения со своей проекцией по любому множеству атрибутов не может изменить исходное отношение, то можно представить исходное отношение m в следующем виде:

$$m = \pi_{TC}(m) \triangleright \triangleleft \pi_{CPV}(m) \triangleright \triangleleft \pi_{TPV}(m).$$

Доказательство завершено.

Замечание.

Отношение $\pi_{TPV}(m)$ может быть получено на основе отношений $\pi_{TC}(m)$ и $\pi_{CPV}(m)$ путем естественного соединения: $\pi_{TPV}(m) = \pi_{TP}(m) \triangleright \triangleleft \pi_{CPV}(m)$ при условии, что в результирующем естественном соединении будет выполнена зависимость $TP \rightarrow V$.

Таким образом, формирование отношения $m : M(T : N, C : \Sigma, P : R, V : N)$ на практике можно представить как последовательность этапов:

1. формирование отношений $r_1 : R_1(T, C)$ и $r_2 : R_2(C, P, V)$ на основе документа описания предметной области;
2. построение отношения $r_3 : R_3(T, P, V)$ по правилу: $r_3 = \pi_{TPV}(r_1 \triangleright \triangleleft r_2)$;
3. проверка выполнимости условий теоремы 1 (проверка корректности).

В свою очередь можно выделить следующие шаги проверки корректности построенных отношений:

1. корректность задания вариантов определения:
 - 1.1. поиск неописанных и неиспользуемых вариантов определения;
 - 1.2. однозначность задания вариантов определения;
2. проверка свойства полусхемы.

Следует подчеркнуть важность порядка выполнения проверок.

Корректность задания вариантов определения.

Отношения r_1 и r_2 могут содержать противоречивую информацию о вариантах использования, поэтому необходимым является условие: $\pi_C(r_1) = \pi_C(r_2)$.

Однозначность задания вариантов определения в силу леммы 2 может быть проверена следующим образом:

$$\forall (f, g \in \pi_C(r_2)) \pi_{PV}(\sigma_{C=f}(r_2)) \neq \pi_{PV}(\sigma_{C=g}(r_2)).$$

Алгоритмы проверки выполнения соответствующих свойств приведены ниже (алгоритмы 1 – 3).

Алгоритм 1. Поиск неописанных и неиспользуемых вариантов определения

```

CheckConformity = proc( $r_1, r_2$ :relation)
  local
     $temp, C_1, C_2$ : TupleSet;
     $str$ : String;
     $C_1 := \pi_c(r_1)$ ; // имена вариантов определения из
отношения  $r_1$ 
     $C_2 := \pi_c(r_2)$ ; // имена вариантов определения из
отношения  $r_2$ 
    if  $C_1 \not\supseteq C_2$  then
       $temp := C_1 \setminus C_2$ ;
      for each  $str$  from  $temp$  do
        print( $str + "$  - неописанный вариант определения");
      end do
    elif  $C_1 \not\subseteq C_2$  then
       $temp := C_2 \setminus C_1$ ;
      for each  $str$  from  $temp$  do
        print( $str + "$  - имя варианта определения," +
          "которое не используется ни " +
          "в одном определении понятия");
      end do
    else
      print("Успешное завершение");
    end if
end proc

```

Алгоритм 2. Проверка однозначности задания вариантов определения

```

UpdateCaseDefinition := proc( $r_1, r_2$ :relation,  $f, g$ : String)
  local
     $Sign, N, R$ : TupleSet;
     $t$ : String;
    // изменение отношения  $r_1$ 
     $N := \pi_T(\sigma_{c=g}(r_1))$ ; // набор понятий с вариантом определения
     $g$ 
     $r_1 := r_1 \setminus \sigma_{c=g}(r_1)$ ; // удаление из  $r_1$  кортежей, содержащих
    вариант // определения  $g$ 
     $R := N \cup \pi_c(\sigma_{c=f}(r_1))$ ;
     $r_1 := r_1 \cup R$ ;
    // изменение отношения  $r_2$ 
     $Sign := \sigma_{c=g}(r_2)$ ;
     $r_2 := r_2 \setminus Sign$ ;
end proc

```

Алгоритм 3. Удаление дублирующихся вариантов определения

```

CheckCaseDefinition := proc ( $r_1, r_2$ :relation)
  local
     $C, Def_1, Def_2$ : TupleSet;
     $f, g$ : String;
     $C := \pi_C(r_2)$ ;
  if  $C \neq \emptyset$  then
    for each  $f, g$  from  $C$  do
       $Def_1 := \pi_{PV}(\sigma_{C=f}(r_2))$ ;
       $Def_2 := \pi_{PV}(\sigma_{C=g}(r_2))$ ;
      if  $Def_1 = Def_2$  then
        UpdateCaseDefinition( $f, g$ );
      end if
    end do
  end if
end proc

```

Для обоснования алгоритмов, реализующих проверку свойств полусхемы нам понадобится следующее

Утверждение 1.

Свойство полусхемы выполнено тогда и только тогда, когда

$$\forall (n \in N) \left| \pi_{TP}(\sigma_{T=n}(r_3)) \right| = \left| \sigma_{T=n}(r_3) \right|.$$

Доказательство.

Очевидно, что для выполнения свойства полусхемы необходимо и достаточно наличие функциональной зависимости $TP \rightarrow V$. Таким образом, нужно доказать, что зависимость $TP \rightarrow V$ выполнена тогда и только тогда, когда

$$\forall (n \in N) \left| \pi_{TP}(\sigma_{T=n}(r_3)) \right| = \left| \sigma_{T=n}(r_3) \right|.$$

Пусть выполнена функциональная зависимость $TP \rightarrow V$, но

$$\forall (n \in N) \left| \pi_{TP}(\sigma_{T=n}(r_3)) \right| < \left| \sigma_{T=n}(r_3) \right|$$

(знак больше невозможен в силу определения проекции). Тогда в отношении r_3 найдутся два кортежа t_1 и t_2 таких, что $t_1 = \langle n, r, n_1 \rangle$, $t_2 = \langle n, r, n_2 \rangle$ и $n_1 \neq n_2$. Но это противоречит условию наличия функциональной зависимости $TP \rightarrow V$, что завершает доказательство в одну сторону.

Обратно. Пусть для всякого $n \in \pi_T(r_3)$ выполнено соотношение

$$\left| \pi_{TP}(\sigma_{T=n}(r_3)) \right| = \left| \sigma_{T=n}(r_3) \right|.$$

Требуется доказать, что отношение r_3 удовлетворяет функциональной зависимости $TP \rightarrow V$. Проведем доказательство от противного. Предположим, что для отношения r_3 функциональная зависимость $TP \rightarrow V$ не выполнена. Эта означает, что

существуют кортежи $t_1 = \langle n, r, n_1 \rangle$, $t_2 = \langle n, r, n_2 \rangle$ с $n_1 \neq n_2$. Следовательно, $|\pi_{TP}(\sigma_{T=n}(r_3))| < |\sigma_{T=n}(r_3)|$, что противоречит условию. Утверждение доказано.

Ниже приведены алгоритмы проверки свойства полусхемы.

Алгоритм 4. Проверка свойства полусхемы для конкретного понятия

```
CheckLURForNotion := proc(r3:relation, n:String)
  local
    count1, count2: Integer;
    count1 :=  $|\sigma_{T=n}(r_3)|$ ;
    count2 :=  $|\pi_{TP}(\sigma_{T=n}(r_3))|$ ;
    if count1 ≠ count2 then
      print(n + "- не удовлетворяет свойству полусхемы");
    end if
end proc
```

Алгоритм 5. Проверка свойства полусхемы для всех понятий

```
CheckLUR := proc(r1: relation)
  local
    T: TupleSet;
    n: String;
    T :=  $\pi_T(r_1)$ ;
    for each n from T do
      CheckLURForNotion(n);
    end do
end proc
```

Представление полусхемы в виде пары отношений упрощает ее анализ, прежде всего это касается двух таких важных этапов, как:

поиск структурных синонимов [1];

поиск некорректно заданных понятий и вариантов определений [1].

Утверждение 2.

Понятия m и n являются структурными синонимами тогда и только тогда, когда

$$\pi_C(\sigma_{T=n}(r_1)) = \pi_C(\sigma_{T=m}(r_1)).$$

Доказательство.

Пусть $F_n = \{f \in M_+(R, N) | (n, f) \in D\}$ и $F_m = \{f \in M_+(R, N) | (m, f) \in D\}$.

По определению понятия n и m являются структурными синонимами тогда и только тогда, когда $F_n = F_m$.

Но множество F_n можно получить из отношения r_1 следующим образом:

$$F_n = \pi_C(\sigma_{T=n}(r_1)).$$

Аналогичным образом получается множество F_m .

Следовательно, сформулированное утверждение является следствием определения структурных синонимов. Доказательство завершено.

На основании этого утверждения может быть построен алгоритм поиска структурных синонимов.

Алгоритм 6. Поиск структурных синонимов
CheckStructSynonyms := proc(r_1 : relation)

```

local
  T, F1, F2: TupleSet;
  m, n: String;
  T1 :=  $\pi_T(r_1)$ ;
  if N ≠ ∅ then
    for each m, n from T do
      F1 :=  $\pi_C(\sigma_{T=n}(r_1))$ ;
      F2 :=  $\pi_C(\sigma_{T=m}(r_1))$ ;
      if F1 = F2 then
        print(m + " и " + n + " - структурные синонимы");
      end if
    end do
  end if
end proc

```

Полусхема, корректная в том смысле, что все определяемые ее понятия имеют не менее одного образца [1], называется схемой. Исходя из этого, может быть построен алгоритм проверки условий схемы, который включает в себя два этапа: поиск плохо определенных понятий и поиск вариантов определения, область значения которых содержит плохо определенные понятия. Вычленение элементов двух построенных множеств из описания полусхемы позволяет утверждать [1], что любое оставшееся понятие имеет хотя бы один образец.

Алгоритм 7. Проверка выполнения условий схемы

```

CheckScheme := proc( $r_1, r_2$ : relation)
local
  N0, N, Sign, Def, F: TupleSet;
  n, f: String;
  N0 :=  $\pi_V(r_2) \setminus \pi_T(r_1)$ ; // множество базовых понятий
  N :=  $\pi_T(r_1) \setminus N_0$ ; // множество небазовых понятий
  for each n from N do
    Sign :=  $\pi_C(\sigma_{T=n}(r_1))$ ; // множество вариантов определения
    // понятия n
    for each f from Sign do
      Def :=  $\pi_V(\sigma_{C=f}(r_2))$ ; // множество значений
      // варианта определения f
      if Def ⊂ N0 then

```

```

    N0 := N0 ∪ {n};
    N := N \ {n};
    break;
end if
end do
end do
// поиск некорректных понятий
if N ≠ ∅ then
    for each n from N do
        print(n + " - некорректное понятие");
    end do
end if
// поиск вариантов определения, область значения
// которых содержит плохо определенные понятия
F := πc(N ▷◁ r2); // множество вариантов определения,
                    // области значений которых содержат
                    // некорректные понятия
for each f from F do
    print (f + "- некорректный вариант определения");
end do
end proc

```

Рассмотрим пример представления концептуально модели, описывающей ER-диаграмму, в виде отношений базы данных. На рис. 1. приведен документ описания предметной области ER-диаграммы.

```

#define er-link =    from-e:entity,
                    to-r: relationship, name:String,
                    cardinality: cardinality_range
#define ea-link =    from-e:entity, to-a: attribute,
                    name: String
#define ra-link =    from-r: relationship,
                    to-a: attribute,
                    name: String
#define functional = functional-value: one-to-many
#define multiple =   multiple-value: many-to-many

entity (name:String);
relationship (name:String);
attribute (name: String, data-type: String)
link(er-link; ea-link; ra-link)
cardinality_range (functional; multiple)

```

Рис. 1. Документ описания предметной области ER-диаграмм

В приведенном документе описания предметной области используются следующие обозначения.

Задание варианта определения:

#define вариант_определения = имя_роли:имя_понятия [, ...]

Определение понятия:

имя_понятия (имя_роли:имя_понятия [, ...] ; [...]) или

имя_понятия (вариант_определения [, ...] ; [...])

На основе приведенных данных построим отношения $r_1(T, C)$ и $r_2(C, P, V)$.

$r_1(T, C)$:

T	C
entity	entity-default
relationship	relationship_default
atribute	atribute_defaulte
link	er-link
link	ea-link
link	ra-link
cardinality_range	functional
cardinality_range	multiple

$r_2(C, P, V)$:

C	P	V
Entity-default	name	String
relationship_default	name	String
atribute_defaulte	name	String
atribute_defaulte	data-type	String
er-link	from-e	entity
er-link	to-r	relationship
er-link	name	String
er-link	cardinality	cardinality_range
ea-link	from-e	entity
ea-link	to-a	atribute
ea-link	name	String
ra-link	from-r	relationship
ra-link	to-a	atribute
ra-link	name	String
functional	functional-value	one-to-many
multiple	multiple-value	many-to-many

На основе двух приведенных отношений легко построить отношение $r_3(T, P, V)$, которое будет являться результатом их естественного соединения.

Используя алгоритмы 1 – 7, приведенные выше, легко убедиться в том, что построенная и представленная в виде набора отношений модель предметной области является схемой.

Таким образом, в настоящей работе предложен метод построения предложенных в [1] полусхем средствами реляционных баз данных.

На основе полученного представления предложены и обоснованы алгоритмы проверки корректности концептуальной модели предметной области, соответствующей полусхеме. Это открывает путь к программной реализации метода в виде CASE-комплекса концептуального моделирования.

ЛИТЕРАТУРА

1. Жолткевич Г.Н., Семенова Т.В. К проблеме формализации концептуального моделирования информационных систем. – Вісник Харківського національного університету імені В.Н. Каразіна, серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління». – № 605, вип. 2. – 2003. – С. 33 – 42.
2. Жолткевич Г.Н., Семенова Т.В. Концептуальное моделирование данных в исследовательских информационных системах средствами реляционных СУБД. – Вестник Херсонского государственного технического университета. – 2002, №15. – С. 75-79.
3. Мейер Д. Теория реляционных баз данных. – М.: «Мир». – 1987. – 608 с.