

УДК 004.23

А.Г. Житарюк, Г.Н. Жолткевич

Харьковский национальный университет им. В.Н. Каразина, Харьков

ПРЕДСТАВЛЕНИЕ ОБРАЗЦОВ КОНЦЕПТОВ ИНФОРМАЦИОННЫХ СИСТЕМ ПОСРЕДСТВОМ АЦИКЛИЧЕСКИХ ГРАФОВ

В работе предложен способ описания образцов концептов информационных систем с помощью ациклических графов. Для хранения граф-образцов концептов был выбран реляционный подход. В рамках этого подхода сформулированы аксиомы для корректного задания образцов.

Ключевые слова: схема данных, ациклический граф, маркированное дерево, образец со связями, свободный образец.

Введение

При разработке программного обеспечения большую роль играет представление данных в виде множества связанных структур. Хорошо спроектированное множество структур данных оптимизирует использование ресурсов, требуемых для выполнения наиболее критичных операций.

С этой точки зрения, полусхему, введенную в работе [2], можно рассматривать как схему данных, тогда как образец концепта как структуру данных, соответствующую схеме.

В связи с этим в работе [2] авторами предложен алгебраический подход к построению образцов концептов полусхем, базирующийся на введенном ими понятии терминальных именуемых нитей. Этот подход предоставил возможность дальнейшего расчленения данных путем наложения условий на экземпляры в терминах многоосновной алгебры. Однако математическая структура, которая моделирует образец понятия, является достаточно формальной, а ее концепция – интуитивно не очевидной. В связи с этим, в работе [1] авторами предложен способ построения моделей образцов понятий полусхем в терминах маркированных деревьев. Такое представление позволило избежать излишнего формализма в определении образцов концептов, а также представить образец в виде специальным образом маркированного дерева.

Поскольку образец как структура данных является моделью схемы данных, задание их посредством маркированного дерева позволяет описывать произвольную схему данных. Для решения реальных задач, необходимо сузить множество всех схем данных до множества допустимых с точки зрения семантики предметной области. Одним из вариантов такого сужения является наложение структурных ограничений на схему данных. В связи с этим, возникает задача построения теории моделирования образцов концептов, которые удовлетворяют заданным структурным ограничениям.

Целью настоящей работы является построение модели хранения образцов концептов информационных систем, для которых заданы структурные ограничения. Задание такого рода ограничений позволит выделить из множества возможных образцов подмножество образцов, которые являются допустимые в семантике данной предметной области. В статье [1] рассматривался способ хранения образцов концептов в виде маркированных деревьев.

В настоящей работе рассматривается модель направленного ациклического графа, как способ задания образцов концептов предметной области, которые удовлетворяют заданным структурным ограничениям.

Модель образцов со связями в виде ориентированного ациклического графа

В работе [1] предложен способ описания образцов концептов с помощью маркированных деревьев. Маркированным деревом авторами было названо дерево, каждая вершина которого помечена либо парой (имя концепта, имя квалификатора), либо именем концепта, в том случае, если он является базовым. Каждое ребро в маркированном дереве помечается именем роли, соответствующей выбранному квалификатору. С помощью маркированного дерева можно задавать произвольные образцы концептов. Такие образцы мы будем называть свободными.

Таким образом, моделями свободных образцов являются специальным образом маркированные деревья.

Пусть задана предсхема (полусхема)

$S = (N, R, D)$ и N_0 – множество ее базовых понятий.

Маркированное дерево

$$T = (V, E, D \cup N_0, R, \text{beg} : E \rightarrow V, \text{end} : E \rightarrow V, m_V : V \rightarrow D \cup N_0, m_E : E \rightarrow R),$$

удовлетворяющее следующим условиям:

1) $m_V(v) \in N_0$ тогда и только тогда, когда v – лист дерева;

2) если $m_V(v) = (n, f) \in D$, то m_E устанавливает взаимно однозначное соответствие между $\{e \in E \mid \text{beg}(e) = v\}$ и $\text{dom}(f)$;

3) если $v, v' \in V, e \in E, \text{beg}(e) = v, \text{end}(e) = v'$, $m_V(v) = (n, f) \in D$, а $m_V(v') = (n', g) \in D$ (для некоторого g) и $m_E(e) = r$, то $(f(r), g) \in D$, или $m_V(v') = n' \in N_0$, то $f(r) = n'$ называется граф-образцом понятия n_t , где $m_V(\text{root}) = f_t$ для некоторого $n_t : (n_t, f_t) \in D$.

Рассмотрим пример.

В качестве примера возьмем предсхему, описывающую ломаную линию.

Пусть дано:

$N = \{\text{'полулиния'}, \text{'сегмент'}, \text{'точка'}, \text{'список'}, \text{'null'}, \text{'Real'}\}$;

$R = \{\text{'голова'}, \text{'хвост'}, \text{'начало'}, \text{'конец'}, \text{'nothing'}, \text{'x'}, \text{'y'}\}$.

Определим пару частичных отображений из R в N следующим образом:

ломанная $\left\{ \begin{array}{l} \text{'голова'} \quad \text{'хвост'} \\ \text{'сегмент'} \quad \text{'список'} \end{array} \right\}$;

состав $\left\{ \begin{array}{l} \text{'начало'} \quad \text{'конец'} \\ \text{'точка'} \quad \text{'точка'} \end{array} \right\}$;

координаты $\left\{ \begin{array}{l} \text{'x'} \quad \text{'y'} \\ \text{'Real'} \quad \text{'Real'} \end{array} \right\}$, empty $\left\{ \begin{array}{l} \text{'nothing'} \\ \text{'null'} \end{array} \right\}$.

Теперь зададим D :

$D = \{(\text{'полулиния'}, \text{ломанная}), (\text{'сегмент'}, \text{состав}), (\text{'точка'}, \text{координаты}), (\text{'список'}, \text{empty}), (\text{'список'}, \text{ломанная})\}$.

Тройка $S = (N, R, D)$ является схемой, моделирующей понятие ломаной линии. Свободный образец ломаной линии представлен на рис. 1.

Для того чтобы данный образец являлся образцом ломаной линии, необходимо выполнения условия – конец каждого сегмента должен совпадать с началом следующего сегмента. При наложении данного условия на образец необходимо связать соответствующие вершины дерева. В связи с этим, пример, рассмотренный в работе, приводит к необходимости построения теории, учитывающей такого рода связи.

Модель маркированного дерева не позволяет отображать структурные связи, поэтому в качестве модели задания образцов концептов мы предлагаем в дальнейшем использовать ациклический граф.

Введем понятия свободных образцов и образцов со связями. Свободными образцами будем называть маркированные деревья, а образцами со связями будем называть подмножество свободных образцов, удовлетворяющих заданным структурным ограничениям.

Для задания образцов концептов со связями необходимо внести изменения в модель маркированного дерева.

Обозначим Q – множество имен квалификаторов.

Ациклическим графом будем называть маркированное дерево T , в котором поставим в соответствие каждой вершине, помеченной именем квалификатора из Q , уникальный идентификатор id . Множество всех идентификаторов для образца концепта обозначим ID .

Введем множество $L : ID \rightarrow Q$ – множества пар (id, q) . $S = (N, R, D)$ – схема предметной области, а N_0 – множество ее базовых понятий.

Тогда модель образца со связями для схемы S будет следующий ациклический граф:

$AG = (V, E, L \cup N_0, R, \text{beg} : E \rightarrow V, \text{end} : E \rightarrow V, m_V : V \rightarrow L \cup N_0, m_E : E \rightarrow R)$,

удовлетворяющий следующим условиям:

1) $m_V(v) \in N_0$ тогда и только тогда, когда v – лист дерева;

2) если $m_V(v) = (id, f) \in L$, то m_E устанавливает взаимно однозначное соответствие между $\{e \in E \mid \text{beg}(e) = v\}$ и $\text{dom}(f)$;

3) если $v, v' \in V, e \in E, \text{beg}(e) = v, \text{end}(e) = v'$, $m_V(v) = (id, f)$, а $m_V(v') = (id', g)$ (для некоторого g) и $m_E(e) = r$, то $(f(r), g) \in D$, или если $m_V(v') = n' \in N_0$, то $f(r) = n'$.

Образец ломаной линии из примера, описанного выше, для которого выполнено ограничение «начало следующего сегмента должно совпадать с концом

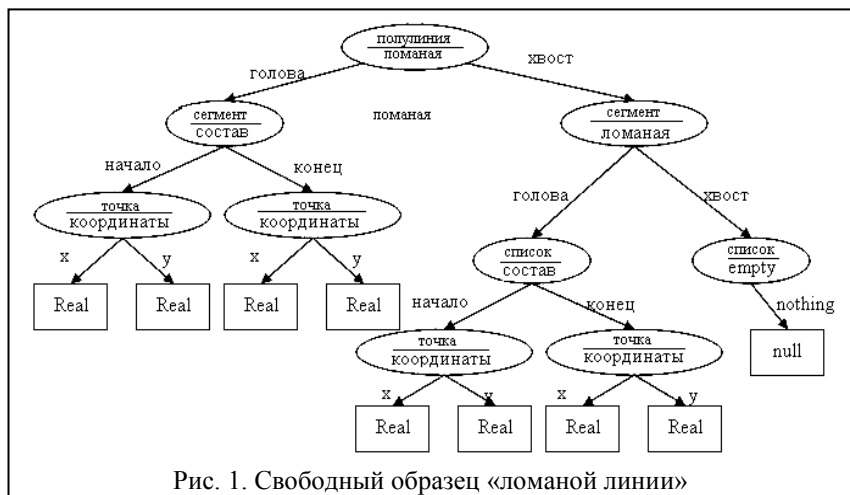


Рис. 1. Свободный образец «ломаной линии»

предыдущего», в виде ациклического графа приведен на рис. 2.

Из рис. 2 видно, что две вершины дерева объединились в одну вершину, помеченную парой (координаты, idC2), что означает – координаты начала следующего сегмента и координаты конца предыдущего – это одни и те же координаты.

Таким образом, с помощью ациклического графа, можно задавать образцы концептов, которые удовлетворяют заданным структурным ограничениям.

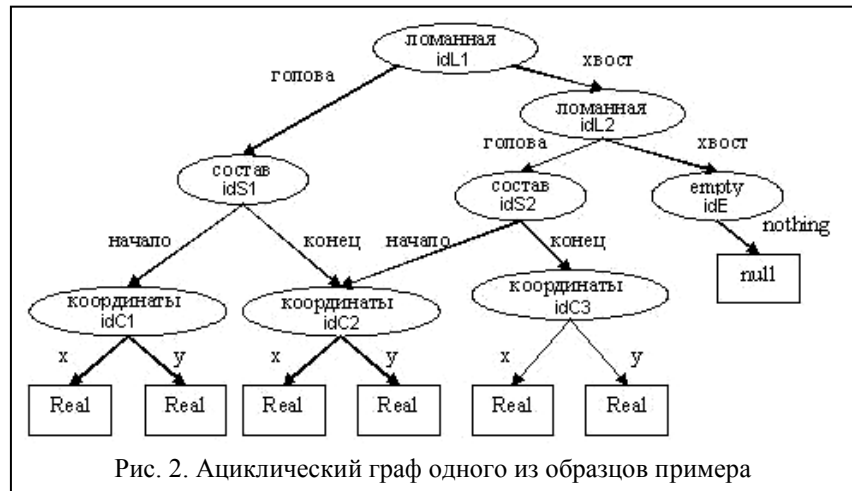


Рис. 2. Ациклический граф одного из образцов примера

Модель ограничения целостности для хранения образцов со связями

В работе [3] показано, что любая предсхема может быть однозначно описана с помощью двух отношений реляционной базы данных следующей структуры:

$$r_1 : R_1(T, C) \text{ и } r_2 : R_2(C, R, V),$$

где атрибут T – имя концепта; атрибут C – имя квалификатора; атрибут R – имя роли; атрибут V – значение квалификатора на роли.

Так как не любая база данных из двух таких отношений, задает некоторую предсхему, в работе [3] сформулированы и доказаны необходимые и достаточные условия, для того чтобы база данных описывала некоторую предсхему, а в работе [4] приведена реализация алгоритмов проверки корректности задания предсхемы средствами реляционной алгебры.

С целью обеспечения вычислительных методов обработки образцов схем, необходима структура данных для их удобного хранения. Представим образец концепта в виде двух отношений реляционной базы данных.

Выбор указанного представления обусловлен технологической зрелостью реляционного подхода, что позволяет при незначительных затратах строить реализации алгоритмов, обеспечивающих проверку корректности моделей.

Отношение IDQ(C, ID) ставит в соответствие каждому имени квалификатора C, множество имен идентификаторов ID, которые его определяют.

Отношение IRD(ID, R, IDD), определяет состав компонентов образца концепта, где ID – имя идентификатора, R – имя роли соответствующего квалификатора, заданного с помощью его идентификатора, и IDD – идентификатор компонента образца, соответствующий роли.

Для наглядности приведем отношения для примера образца ломаной линии.

Отношение IDQ

C	ID
ломанная	idL1
ломанная	idL2
состав	idS1
состав	idS2
координаты	idC1
координаты	idC2
координаты	idC3
empty	idE

Отношение IRD

ID	R	IDD
idL1	голова	idS1
idL1	хвост	idL2
idL2	голова	idS2
idL2	хвост	idE
idS1	начало	idC1
idS1	конец	idC2
idS2	начало	idC2
idS2	конец	idC3
idC1	x	Real
idC1	y	Real
idC2	x	Real
idC2	y	Real
idC3	x	Real
idC3	y	Real
idE	nothing	null

Однако, не любые два отношения с приведенными выше схемами, будут описывать образец концепта предметной области. Необходимо сформулировать ряд аксиом для однозначного восстановления образца концепта, по фрагменту базы данных, его описывающему.

Аксиоматика

Пусть задана предсхема, с помощью отношений $r_1 : R_1(T, C)$ и $r_2 : R_2(C, R, V)$, и образец, описанный с помощью отношений IDQ(C, ID) и IRD(ID, R, IDD). Для того, чтобы образец соответствовал предсхеме, необходимо выполнение следующих аксиом:

Аксиома 1. Для отношения IDQ должна выполняться функциональная зависимость $ID \rightarrow C$, т.е. каждый идентификатор однозначно определяет квалификатор

Аксиома 2. $\pi_C(IDQ) \subset \pi_C(r_1)$, т.е. множество имен квалификаторов для образца является подмножеством всех возможных квалификаторов предсхемы.

Аксиома 3. Для каждого квалификатора в отношении IRD должны быть представлены все его компоненты

Аксиома 4. В описании каждого идентификатора не должен присутствовать он сам в качестве одного из компонентов.

Аксиома 5. Образец должен иметь точно такую же структуру, как и его предсхема

Для выполнения аксиом 3 – 5 ниже приведены процедуры, описанные на языке реляционной алгебры.

Процедура проверки аксиомы 3:

```
begin
for each q ∈ πC(IDQ)
  temp = πID(σC=q(IDQ))
  for each id ∈ temp
    if πR(σID=id(IRD)) ≠ πC(σC=q(r2))
      sample incorrect
      break;
    end if
  end for
end for
sample correct
end
```

Процедура проверки аксиомы 4:

```
begin
for each id ∈ πC(r1)
  temp = πV(σC=id(r2))
  if σV=id(temp) ≠ Π
    sample incorrect
    break;
  end if
  else
    for each t ∈ temp
      temp' = πV(σC=t(V)(r2))
      temp = temp ∪ temp' \ t
      if σV=id(temp) ≠ Π
        sample incorrect
        break;
      end if
    end for
  end else
end for
sample correct
end
```

Алгоритм проверки аксиомы 5.

Построим для образца вспомогательные отношения r_{temp1} и r_{temp2} :

$$r_{temp1} = \pi_{C,R,IDD}(IDQ \triangleright \triangleleft IRD);$$

$$r_{temp2} = \pi_{C,R,IDD,Q}(r_{temp1}[IDD = ID] \rho_{C=Q}(IDQ)).$$

Также построим вспомогательное отношение и для схемы.

$$r_{schema} = \pi_{C,R,V,IDD}(\rho_{C=IDD}(r_1[T = V]r_2)).$$

Если выполнено

$$\rho_{Q=IDD}(\pi_{C,R,IDD}(r_{temp2})) \cup r_{temp1} \setminus \pi_{C,R,IDD}(r_{temp2}) \subset \subset (\pi_{C,R,IDD}(r_{schema}) \cup \rho_{V=IDD}(r_2 \setminus \pi_{C,R,V}(r_{schema}))),$$

тогда образец не противоречит схеме.

Выводы

Таким образом, в статье предложена модель ациклического графа, как способ задания образцов со связями, а также модель реляционной базы данных для ее хранения. Для предложенной модели был сформулирован ряд аксиом. Для каждой аксиомы приведен алгоритм проверки, описанный на языке реляционной алгебры.

Описания образцов концептов предметной области посредством ациклических графов позволит выделить среди множества всех образцов, подмножество образцов, удовлетворяющих структурным ограничениям.

Список литературы

1. Жолткевич Г.Н. О возможности представления образцов понятий полусхем маркированными деревьями / Г.Н. Жолткевич, Ахмад Юсеф Ибрахим Ибрахим. // Системи обробки інформації: зб. наук. пр. – Х.: ХУПС, 2006. – Вип. 2 (51). – С. 20-26.
2. Жолткевич Г.Н. К проблеме формализации концептуального моделирования информационных систем / Г.Н. Жолткевич, Т.В. Семенова // Вісник Харк. нац. ун-та. Серія „Мат. моделювання. Інформ. технології. Автоматизовані системи управління”. – 2003. – № 605 (2). – С. 33-42.
3. Жолткевич Г.Н. Представление полусхем предметных областей информационных систем средствами реляционных баз данных / Г.Н. Жолткевич, Т.В. Семенова, К.А. Федорченко // Вісник Харк. нац. ун-та. Серія „Математичне моделювання. Інформаційні технології. Автоматизовані системи управління”. – 2004. – № 629 (3). – С. 11-24.
4. Жолткевич Г.Н. Проверка корректности спецификации концептуальной модели предметной области средствами реляционной алгебры / Г.Н. Жолткевич, К.А. Федорченко // Вестник Херсонского нац. техн. ун-та. – 2005. – № 22. – С. 138-142.

Поступила в редколлегию 17.09.2010

Рецензент: д-р техн. наук, проф. В.А. Филатов, Харьковский национальный университет радиоэлектроники, Харьков

ПРЕДСТАВЛЕННЯ ЗРАЗКІВ КОНЦЕПТІВ ІНФОРМАЦІЙНИХ СИСТЕМ ЗА ДОПОМОГОЮ АЦИКЛІЧНИХ ГРАФІВ

А.Г. Житарюк, Г.М. Жолткевич

В роботі запропонований засіб описання зразків концептів інформаційних систем за допомогою ациклических графів. Для збереження граф-зразків концептів був обраний реляційний підхід. У межах цього підходу сформульовані аксіоми для визначення зразків.

Ключові слова: ациклический граф, маркіроване дерево, зразок з зв'язками, вільний зразок.

REPRESENTATION OF THE CONCEPT SAMPLES OF THE INFORMATION SYSTEMS BY MEANS OF ACYCLIC GRAPH

A.G. Zhytaruk, G.M. Zholtkevych

Method of the Representation of the concept samples of the information systems by means of the acyclic graph is considered. The relational approach for storing of the graph-samples is chosen. A number of axioms within the bounds of this approach are formulated.

Keywords: *acyclic graph, marking tree, sample with relation, free sample.*