

В. А. Овчинников, Г. С. Иванова

**МЕТОДИКА ФОРМАЛЬНОГО СИНТЕЗА
КОМБИНИРОВАННЫХ СТРУКТУР ДАННЫХ
ДЛЯ ПРЕДСТАВЛЕНИЯ ГРАФОВ**

На примере организации хранения множества ребер гиперграфа и их образов относительно предиката инцидентности рассмотрена методика синтеза комбинированных многоуровневых структур данных. Математическими моделями базовых и производных структур данных являются ориентированные графы. Модель синтезированной структуры формируется в результате выполнения операции объединения модели исходной структуры и модели отношений, обеспечивающих эффективную реализацию заданных операций.

E-mail: vaovchinnikov@gmail.com, gsivanova@gmail.com

Ключевые слова: графы, математические модели, структуры данных, формальный синтез.

Формальный синтез комбинированных структур базируется на анализе данных, отношений между ними и набора выполняемых операций. Целью синтеза является получение такой структуры, которая обеспечивала бы высокую эффективность выполнения заданных над данными операций.

Возможность формального синтеза обеспечивается наличием библиотеки моделей базовых и производных структур данных, а также процедур отношений и предикатов (биективного, сюръективного и инъективного отношений, предиката принадлежности элементов подмножества множеству и т. д.). При автоматизированном синтезе пользователю необходимо задать имена, размеры множеств, отношения между ними и имена моделей базовых или производных структур для их хранения.

Рассмотрим механизм такого синтеза на следующем примере. Пусть необходимо построить структуру для представления множества ребер и их образов гиперграфа, заданного в форме $H(X, \langle U, W \rangle, GX, GU)$. Над гиперграфом выполняются операции поиска ребер с максимальным весом, удаления вершин и «пустых» ребер, т. е. тех, у которых $|Gi| = 0$. Количество ребер $|U|$ после удаления «пустых» может быть значительно меньшим, чем до удаления, которое обозначим как $|U_{исх}|$. При удалении вершин и «пустых» ребер порядок их записи необходимо сохранять. Все операции следует выполнять с минимально возможной вычислительной сложностью. Дальнейшие оценки даются в количестве операций сравнения для худшего случая.

С учетом сказанного выше множество ребер гиперграфа U целесообразно представить в виде двусвязного списка L_D [1]. Также в виде двусвязных списков $\{L_j / j = 1, m\}$ представим множества вершин $X_j = \Gamma u_j$, инцидентных каждому ребру u_j (рис. 1).

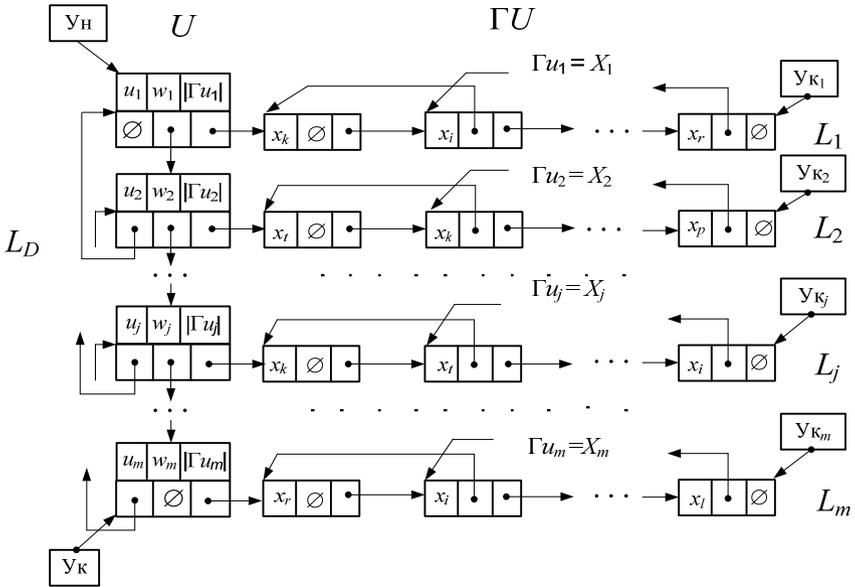


Рис. 1. Двусвязный список двусвязных списков для хранения множеств $\langle U, W, |\Gamma U| \rangle$ и ΓU

Обозначим через x_i вершину, удаляемую на i -м шаге алгоритма. Поиск удаляемой вершины x_i во множестве ΓU подразумевает определение ребер $U_i = \Gamma x_i$, инцидентных этой вершине. Определение множества ребер U_i выполняется по множеству образов ΓX , т. е. не относится к синтезируемой структуре.

На i -м шаге алгоритма над множеством ребер U и множеством подмножеств их образов ΓU будут выполняться следующие операции:

- определение в множестве U ребра u_k с максимальным весом;
- поиск каждого ребра $u_j \in U_i$ в множестве U ;
- поиск удаляемой вершины x_i в образах Γu_j ;
- удаление этой вершины;
- определение «пустых» ребер;
- поиск «пустого» ребра в множестве U ;
- удаление «пустого» ребра.

Для определения «пустых» ребер необходимо знать мощность множества Γu_j каждого ребра $u_j \in U_i$. Вычислительная сложность та-

кой операции $Q = Ar$, которую можно выполнять с вычислительной сложностью $Q = \rho$, если каждому ребру $u_j \in U$ поставить в однозначное соответствие мощность его образа $|Gu_j|$ и уменьшать эту мощность на 1, если удаляемая вершина $x_i \in Gu_j$. Таким образом, каждое ребро гиперграфа будет задаваться тройкой $\langle u_j, w_j, |Gu_j| \rangle$.

Моделью исходной двухуровневой структуры двусвязный список двусвязных списков является граф $\bar{G}_{DL} (\{Z_y, Z_{\Delta D}, Z_Y, Z_{\Delta L}, Y_D, Y_L\}, FZ_D, FZ_L)$. Шаблон этого графа должен находиться в библиотеке моделей базовых и производных структур данных (модели базовых и некоторых комбинированных и двухуровневых структур данных рассмотрены в работах [2, 3]).

Для синтезируемой структуры необходимо задать:

- вид структуры и имя ее модели;
- множество кортежей $\{\langle u_j, w_j, |Gu_j| \rangle / j = 1, |U_{исх}|\}$, которые являются значениями элементов двусвязного списка L_D , и мощность $m = |U_{исх}|$ этого множества ($U_{исх}$ – исходное множество ребер гиперграфа);
- имена множеств $\{X_j / j = 1, |U_{исх}|\}$, значение элементов которых принимает первое поле записи элементов двусвязных списков $\{L_j / j = 1, |U_{исх}|\}$, и мощности $\{|X_j| / j = 1, |U_{исх}|\}$ этих множеств.

В результате настройки шаблона получим граф $\bar{G}_{DL} (\{Z_D, Z_L\}, \{Y_D, Y_L\}, FZ_D, FZ_L)$, в котором (рис. 2)

- $Z_D = \{Z_{yD}, Z_{\Delta D}\}$;
- $Z_{yD} = \{z_{y.н}, z_{y.к}\}$, $z_{y.н} \leftrightarrow a_{y.н}$, $z_{y.к} \leftrightarrow a_{y.к}$ – вершины, сопоставленные адресам указателей начала и конца списка L_D соответственно;
- $Z_{\Delta D} = \{zd_j / j = 1, |U_{исх}|\}$, $Z_{\Delta D} \leftrightarrow A_{\Delta D}$, где $A_{\Delta D}$ – адреса элементов двусвязного списка L_D , $zd_j \leftrightarrow a_{\Delta Dj}$;
- $Z_L = \{Z_{yL}, Z_{\Delta L}\}$;
- $Z_{yL} = \{Z_j / j = 1, |U_{исх}|\}$, $Z_j = \{z_{y.нj}, z_{y.кj}\}$, $z_{y.нj} \leftrightarrow a_{y.нj}$, $z_{y.кj} \leftrightarrow a_{y.кj}$ – вершины, сопоставленные адресам указателей начала и конца списков $\{L_j\}$;
- $Z_{\Delta L} = \{Z_{\Delta Lj} / j = 1, |U_{исх}|\}$, $Z_{\Delta Lj} \leftrightarrow A_{\Delta Lj}$, где $A_{\Delta Lj} = \{aj_i / i = 1, |X_j|\}$ – адреса элементов списка L_j , $Z_{\Delta Lj} = \{zj_t / t = 1, |X_j|\}$, $zj_t \leftrightarrow aj_t$, $|X_j| = Gu_j$;
- $Y_D = \{yd_j / j = 1, |U_{исх}|\}$, $yd_j = \langle yd'_j, z_{y.нj} \rangle$, $yd'_j \leftrightarrow \langle u_j, w_j, |Gu_j| \rangle$;
- $Y_L = \{Y_{Lj} / j = 1, |U_{исх}|\}$, $Y_{Lj} \leftrightarrow X_j$, $X_j = Gu_j$, $Y_{Lj} = \{yj_t / t = 1, |X_j|\}$, $yj_t \leftrightarrow x_i$;
- $FZ_D = \{FZ_{yD}, FZ_{\Delta D}\}$, $Fz_{y.н} = \{zd_1\}$, $Fz_{y.к} = \{zd_m\}$;
- $FZ_{\Delta D} = \{Fzd_j / j = 1, |U_{исх}|\}$, $Fzd_j = \langle yd_j, zd_{j-1}, zd_{j+1} \rangle$;
- $FZ_L = \{FZ_{yL}, FZ_{\Delta L}\}$, $FZ_{yL} = \{Fz_{y.нj}, Fz_{y.кj}\}$, $Fz_{y.нj} = \{zj_1\}$, $Fz_{y.кj} = \{zj_m\}$;
- $FZ_{\Delta L} = \{FZ_{\Delta Lj} / j = 1, |U_{исх}|\}$, $FZ_{\Delta Lj} = \{Fzj_t / t = 1, |X_j|\}$, $Fzj_t = \langle yj_t, zj_{t-1}, zj_{t+1} \rangle$.

вершины x_i в множестве $A_{\exists L} = \{A_{\exists j} / j = 1, |U_{\text{исх}}|\}$ адресов элементов списков $\{L_j\}$. Соответственно, для всего множества вершин X необходимо иметь множество списков $S = \{S_i / i = 1, |X_{\text{исх}}|\}$, что обеспечит прямой доступ к каждому элементу x_i по всем $\{\Gamma u_j\}$.

Моделью каждого односвязного списка адресов S_i является граф $\vec{G}_{S_i}(Z_i, Y_i, FZ_i)$, в котором:

- $Z_i = \{z_{y, n S_i}, Z_{\exists i}\}$;
- $z_{y, n S_i} \ll a_{y, n S_i}, a_{y, n S_i}$ – указатель начала каждого списка S_i ;
- $Z_{\exists i} = \{z_{ik} / k = 1, |\Gamma x_i|\}$, $z_{ik} = z_{jr}$, если $x_i \ll y_{jr}$ и y_{jr} является первым элементом кортежа $\langle y_{jr}, z_{j_{r-1}}, z_{j_{r+1}} \rangle = Fz_{jr}$, $Z_{\exists i} \subset Z_{\exists L}$, $Z_{\exists L} = \{Z_{\exists Lj} / j = 1, |X_{\text{исх}}|\}$, т. е. $Z_{\exists i}$ – это подмножество вершин, сопоставленных адресам тех элементов множества двусвязных списков $\{L_j\}$, значениями которых является вершина x_i ;
- $Y_i = Z_{\exists i} \setminus z_{i1}$;
- $FZ_i = \{Fz_{y, n S_i}, FZ_{\exists i}\}$, $Fz_{y, n S_i} = \{z_{i1}\}$;
- $FZ_{\exists i} = \{z_{ik} / k = 1, p\}$, $p = |\Gamma x_i|$, $Fz_{ik} = \{z_{i_{k+1}}\}$ для $k = 1, p - 1$ и $Fz_p = \emptyset$.

Обозначим множество графов \vec{G}_{S_i} как $\vec{G}_S(Z_S, Y_S, FZ_S) = \{\vec{G}_{S_i}(Z_i, Y_i, FZ_i) / i = 1, |X_{\text{исх}}|\}$, где $Z_S = \{Z_i / i = 1, |X_{\text{исх}}|\}$, $FZ_S = \{FZ_i / i = 1, |X_{\text{исх}}|\}$.

Объединив по вершинам множеств $Z_{\exists L}$ и $\{Z_{\exists i}\}$ графы \vec{G}_{DLV} и $\vec{G}_S - \vec{G}_{DLVS}(Z2, Y1, FZ2) = \vec{G}_{DLV}(Z1, Y1, FZ1) \cup \vec{G}_S(Z_S, Y_S, FZ_S)$, где $Z2 = \{Z1, Z_{YS}\}$, $Z_{YS} = \{z_{y, n S_i} / i = 1, |X_{\text{исх}}|\}$, $FZ2 = \{FZ1, FZ_{YS}\}$, $FZ_{YS} = \{Fz_{y, n S_i} / i = 1, |X_{\text{исх}}|\}$ и $Fz_{jt} = \langle y_{jt}, z_{j_{t-1}}, z_{j_{t+1}}, z_{i_{k+1}} \rangle$, получим модель трехсвязного списка (рис. 4).

Однако для удаления вершины x_i необходимо найти соответствующий список S_i . Таким образом, поиск и удаление вершины будет выполняться с вычислительной сложностью $Q = n \rho$. Для исключения поиска списка S_i зададим отношение D в трактовке «координата элемента x_i в векторной структуре записи множества $X_{\text{исх}}$ соответствует координате элемента», определяющего начало списка S_i в записи множества S . Очевидно, что в полученной структуре элементы соответствующего вектора прямого доступа, назовем его P , будут иметь значение адресов $a_{y, n i}$ указателей начала каждого списка S_i .

Моделью вектора прямого доступа P является граф: $\vec{G}_P(Z_P, Y_P, FZ_P)$, в котором:

- $Z_P = \{z_{\exists P}, Z_{\exists P}\}$, $z_{\exists P} \ll a_{\exists P}, a_{\exists P}$ – адрес базы вектора P ;
- $Z_{\exists P} = \{z_{pi} / i = 1, |U_{\text{исх}}|\}$, $Z_{\exists P} \ll A_{\exists P}, A_{\exists P}$ – адреса элементов вектора P ;
- $Y_P = \{y_{pi} / i = 1, |X_{\text{исх}}|\}$, $y_{pi} = z_{y, n S_i}$ – если x_i не удален из множества X , и $y_{pi} = 0$ в противном случае;
- $FZ_P = \{F_1 z_{\exists P}, F_2 Z_{\exists P}, F_3 Z_{\exists P}\}$, $F_1 z_{\exists P} = Z_{\exists P}$, $F_2 Z_{\exists P} = \{F_2 z_{pi} / i = 1, |X_{\text{исх}}|\}$, $F_2 z_{pi} = Z_{\exists P} \setminus z_{pi}$, $F_3 Z_{\exists P} = \{F_3 z_{pi} / i = 1, |X_{\text{исх}}|\}$, $F_3 z_{pi} = y_{pi}$.

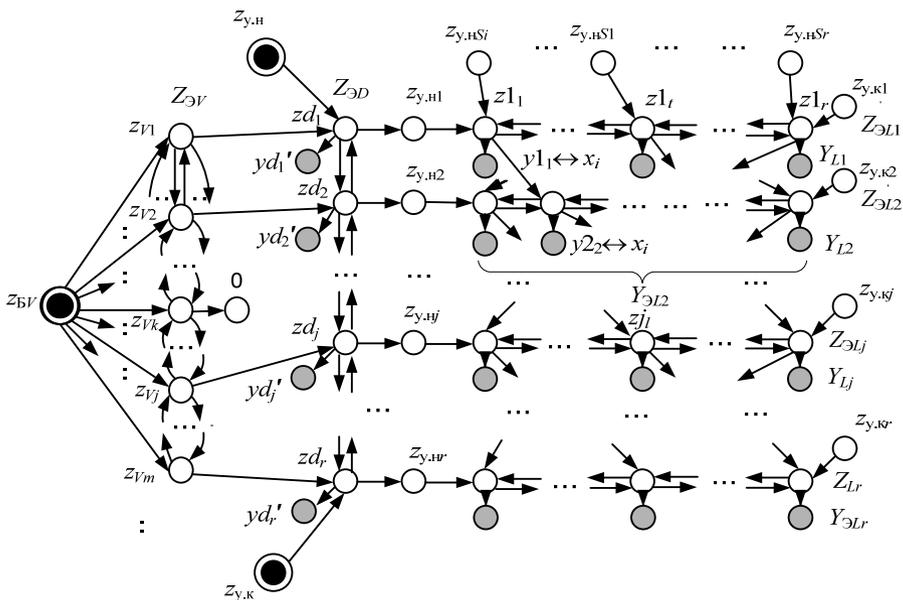


Рис. 4. Модель двухуровневой структуры трехсвязный список с вектором прямого доступа V к вершинам модели списка L_D :

$$y d'_j \leftrightarrow \langle u_j, w_j, |G_{u_j}| \rangle, r = |U|$$

В результате объединения по вершинам u_{p_i} и $z_{y,h S_i}$ графов \vec{G}_{DLVS} и \vec{G}_P

$$\vec{G}_{D-P}(Z, Y, FZ) = \vec{G}_{DLVS}(Z2, Y1, FZ2) \cup \vec{G}_P(Z_P, Y_P, FZ_P),$$

получим модель комбинированной структуры – трехсвязный список с векторами прямого доступа к элементам списка L_D и к спискам S_i . Здесь $Z = \{Z2, Z_P\}$, $Y = \{Y1, Y_P\}$, $FZ = \{FZ2, FZ_P\}$ (рис. 5).

Теперь удаление вершины x_i заключается в обнулении $Y_{P_i} \in Y_P$ графа \vec{G}_P , удалении вершины $Z_{\Delta i} \subset Z_{\Delta L}$, т. е. графа \vec{G}_{S_i} , и выполняется с вычислительной сложностью, равной ρ .

Таким образом, формальный синтез структуры производится операцией объединения рассмотренных выше графов:

$$\vec{G}_C(Z, Y, FZ) = \vec{G}_{DL}(\{Z_D, Z_L\}, \{Y_D, Y_L\}, FZ_D, FZ_L) \cup \vec{G}_V(Z_V, Y_V, FZ_V) \cup \vec{G}_S(Z_S, Y_S, FZ_S) \cup \vec{G}_P(Z_P, Y_P, FZ_P).$$

Модели рассмотренных структур, содержащие информацию, необходимую для оценки емкостной и вычислительной сложности операций над ними имеют вид:

- каждого двусвязного списка данных L_j

$$\vec{G}_{L_j}(\{\langle \{z_{y,hj}, z_{y,kj}\}, v_A \rangle, \langle Z_{\Delta Lj}, Q, v_A \rangle\}, \langle Y_{Lj}, v_3 \rangle, F\{z_{y,hj}, z_{y,kj}\}, FZ_{\Delta Lj}),$$

$$j = 1, m;$$

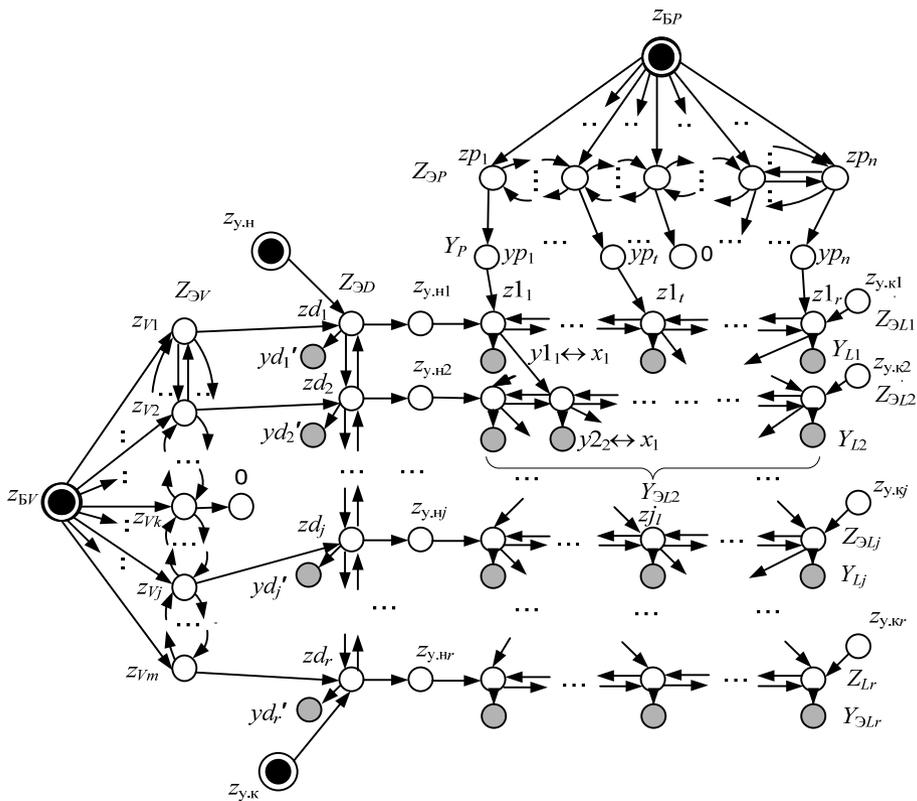


Рис. 5. Модель двухуровневой структуры трехсвязный список с векторами прямого доступа к вершинам модели списка L_D и вершинам – указателям начала списков $\{S_i\}$:

$$y d'_j \leftrightarrow \langle u_j, w_j, | \Gamma u_j | \rangle, r = |U|, y p_i = z_{y,н Si}$$

- двусвязного списка данных LD

$$\vec{G}_{LD} (\{ \langle \langle Z_{yD}, v_A \rangle, \langle Z_{ЭD}, Q, v_A \rangle \rangle, \langle Y_D, v_3, v_A \rangle \rangle, FZ_D \});$$

- вектора адресов прямого доступа V

$$\vec{G}_V (\{ \langle \langle z_{БВ}, v_A \rangle, \langle Z_{ЭВ}, Q \rangle \rangle, \langle Y_V, v_A \rangle \rangle, FZ_V \});$$

- вектора адресов прямого доступа P

$$\vec{G}_P (\{ \langle \langle z_{БP}, v_A \rangle, \langle Z_{ЭP}, Q \rangle \rangle, \langle Y_P, v_A \rangle \rangle, FZ_P \});$$

- каждого односвязного списка адресов S_i

$$\vec{G}_{S_i} (\{ \langle \langle z_{y,н Si}, v_A \rangle, \langle Z_{Эi}, Q, v_A \rangle \rangle, FZ_i \}), i = 1, n.$$

Емкостная сложность комбинированной структуры определяется как сумма весов v вершин объединенного графа G_c . С учетом объединения указанных выше множеств получим

$$V_c = v_A (3 A m + 5 m + n + 4) + v_3 m (A + 3),$$

где $A = | \Gamma u |$.

Полученная синтезированная двухуровневая структура представляет собой трехсвязный список с векторами прямого доступа к элементам списка L_D и указателям начала списков $\{S_i\}$ (рис. 6). Вычислительная сложность выполнения основных операций доступа к элементам первого и второго уровня над этой структурой данных благодаря реализации в структуре дополнительных отношений составит:

- 1) в множестве $\langle U, W, |\Gamma U| \rangle$: поиск элемента
 - по номеру $q_i = 1$,
 - по значению (в том числе максимального и минимального) $q_{vol} = m / 2$;
- 2) в множествах Γu_j :
 - удаление элемента x_k (через двусвязный список S_i) $q_{del} = 1$.
 - операция $\forall u_j \in \Gamma x_k : \Gamma u_j := \Gamma u_j \setminus x_k$.

В результате операция будет выполняться с функциональной вычислительной сложностью ρ , в то время как в структуре, показанной на рис. 1, со сложностью, равной $\rho(m + A)$.

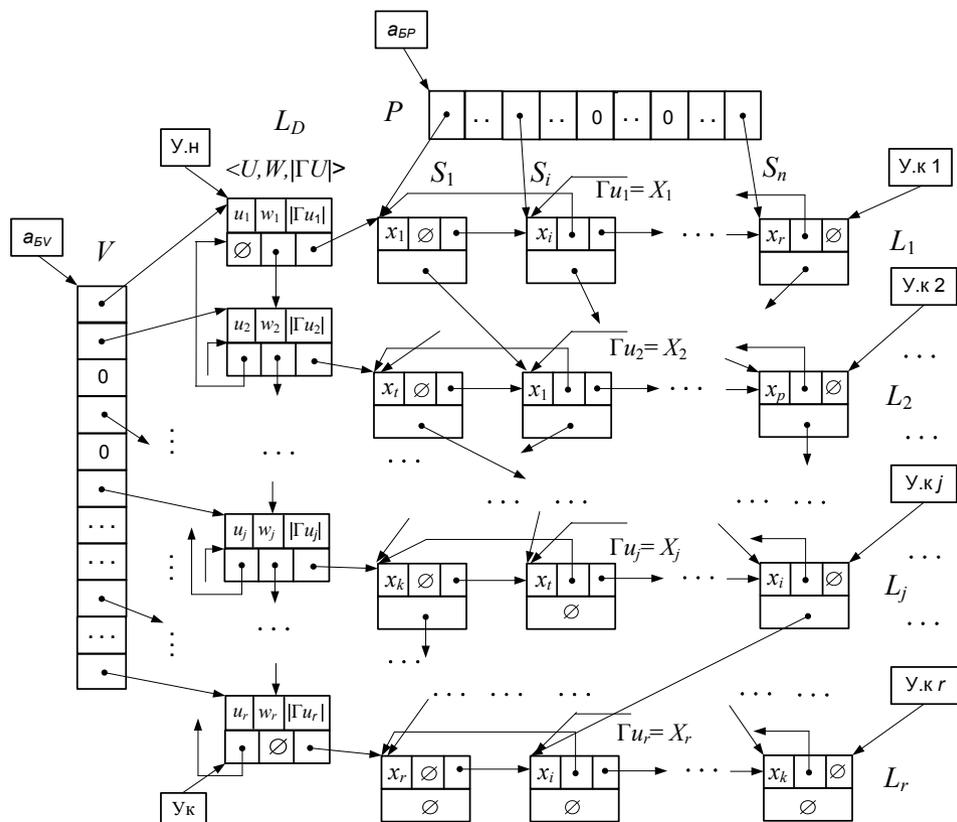


Рис. 6. Синтезированная двухуровневая структура – трехсвязный список с векторами прямого доступа к элементам списка L_D и указателям начала списков $\{S_i\}$

По постановке задачи синтеза структур данных для представления объектов задач рассматриваемого класса на основании выполненного анализа можно сформулировать следующие положения:

1) объект задачи структурного анализа или синтеза (граф) задается аналитически совокупностью множеств и множеств подмножеств. (При матричном задании задача синтеза эффективных структур данных не стоит.);

2) между элементами указанных множеств и подмножествами существуют отношения соответствия, определяемые предикатами инцидентности или смежности, которые востребованы для данного алгоритма. Эти отношения являются взаимно однозначными и должны быть реализованы в производной структуре данных для аналитического представления графа;

3) структуры данных для представления множеств и подмножеств выбирают в соответствии с требованием минимальной вычислительной сложности выполнения комплекса операций, заданных над этими множествами в алгоритме, с учетом ограничения на емкостную сложность представления объекта (вектор, односвязный или двусвязный список);

4) необходимость эффективного выполнения на полученном представлении объекта операций, связанных с преобразованием соответствующих иерархических структур данных, обуславливает появление дополнительных отношений и реализующих их структур;

5) каждая дополнительная структура требует затрат памяти на ее организацию.

Исходными данными для синтеза структуры, представляющей объект задачи структурного анализа или синтеза являются:

- совокупность множеств, представляющих объект и основные отношения между этими множествами и их элементами, а также размер элемента данных каждого множества;

- множество базовых и производных структур данных для представления множеств и их подмножеств;

- множество операций, выполняемых над объектом, и описания этих операций на языке уровня множеств, что позволит определить множество операций над данными структуры;

- количество повторений каждой операции над объектом в алгоритме;

- совокупность дополнительных отношений между компонентами данных, реализация которых позволяет задать альтернативные варианты выполнения операций над синтезируемой структурой данных с меньшей вычислительной сложностью;

- совокупность структур, реализующих дополнительные отношения в основной структуре данных;

- допустимый объем памяти, который может занимать синтезируемая структура.

Наличие указанных данных позволяет формировать множество вариантов модели структуры данных, которые получают объединением тех или иных моделей базовых, производных и дополнительных структур, и определить их характеристики. Из полученного множества необходимо выбрать структуру, которая для заданного набора операций обеспечивает минимальную вычислительную сложность, и емкостная сложность которой не превышает допустимого значения.

СПИСОК ЛИТЕРАТУРЫ

1. О в ч и н н и к о в В. А. Алгоритмизация комбинаторно-оптимизационных задач при проектировании ЭВМ и систем. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2001. – 288 с.
2. И в а н о в а Г. С. Математические модели структур данных // Информационные технологии. – 2006. – № 9. – С. 44–52.
3. И в а н о в а Г. С. Методология и средства разработки алгоритмов решения задач анализа и синтеза структур программного обеспечения и устройств вычислительной техники: Дис. ... д-ра техн. наук: 05.13.11. – М., 2007.

Статья поступила в редакцию 14.05.2012