INFORMATICA, 1993, Vol.4, No.3–4, 351–359

CONCEPTUAL AND RELATIONAL SCHEMES OF ENTITIES

Bronius PARADAUSKAS

Department of Information Systems
Kaunas University of Technology
3028 Kaunas, Studentų St. 50–108, Lithuania

Abstract. The universal structural type of information units-entities is presented the paper. This approach can be recommended for data base modelling at the conceptual and relational levels.

Key words: entities, conceptual scheme, relational scheme, structural type, commutational diagram, compounded functional dependence.

1. Introduction. In order to create reliable programs of informational tasks which are positively evaluated by the end-users of the results, it is essential firstly to analyse the problem area in which the users work, and then to present an expert evaluation of it.

For a long time specialists of informatics build and practically implement conceptual models (CM). Those CM are used as tool for integration of informational needs of the end-users of information systems (IS) and for the administration of expert knowledge about the problem area. CM serves as a tool of arranging the interface between the end users and the programmers. In some cases the attempt are made creating interpreters or generators for automatic task programming by means of CM.

In the process of accumulating and representing knowledge about the problem area informatics specialists and experts manipulate entities the name of which are included in CM. So, conceptual models represent the regularities of the problem domain and semantical data properties to be simulated. Special expert systems and languages are known for these purposes; see reviews in (Daniel, 1990; Chen, 1976; Jasiukevičius, 1986). The entities used by end users, characterize denotants that are the objects of a computerized information system. References to denotants are expressed by means of entities properties. Cause-effect phenomena of the problem area, represented with generalization and abstraction, causes taxonomic connections between the entities. Connections of this kind are the paths of transition from more general entities to the more detailed ones. Formally these connections represent partial ordering: they are transitive, reflexive and antisymmetric. The approach used in this paper allows to chose the formal apparatus for ordering the universum of entities names, and to define structural type. It follows the approach used in (Jasiukevičius, 1986). Now we will assume that entities of a different level of detail have their own names which may be used as elements of CM or as data instances at the level of database realization.

2. Denotational semantics for structural types of object names. Analysis of the structural entities name sets has been based on the initial assertion that name universum D is a complete partially ordered set (CPOS). For this purpose the least element $\perp$ such that $\forall x \in D\ [\perp \sqsubseteq x]$ is introduced and there is a least upper bound $\sqcup X \in D$ for any directed subset $X \subseteq D$. The approximation symbol $\sqsubseteq$ indicates the partial ordering of the elements of a directed subset $X \subseteq D$. The subset $X \subseteq D$ is considered to be directed because it is not empty and

$$\forall x, y \in X \quad \exists z \in X\ [x \sqsubseteq z,\ y \sqsubseteq z].$$

Let us introduce two operators r and p on the set D as CPOS:

- r operator of constructing positioned subsets of D;
- p operator of constructing a functional set on D.

The operator $r:\ D \to r(D)$ is used in developing a new set and extends the domain of the definition to

$$r(D) = D + D^2 + \cdots + D^{|D|}. \tag{2.1}$$

According to (2.1) the set $r(D)$ is obtained by union of sets D^n, $n = 1, 2, \dots |D|$ with artificially introduced least elements

$$\bot_{D^n} = \langle \underbrace{\bot_D, \bot_D, \dots, \bot_D}_{n} \rangle; \quad n = 1, 2, \dots, |D|. \tag{2.2}$$

The ordering $\sqsubseteq r(D)$ on the set $r(D)$ is induced by $\sqsubseteq_D$ on CPOS D:

$$\langle x_1, x_2, \dots x_n \rangle \sqsubseteq_{r(D)} \langle y_1, y_2, \dots, y_n \rangle, \tag{2.3}$$

only when $x_i \sqsubseteq_D y_i$, $i = 1, 2, \dots, n$; $n \leqslant |D|$.

The upper bound for the two elements

$$\langle x_1, x_2, \dots, x_n \rangle \in X \quad \text{and} \quad \langle y_1, y_2, \dots, y_n \rangle \in X, \quad X \subseteq D^n$$

is calculated in the following way:

$$\langle x_1, x_2, \dots, x_n \rangle \overset{D^n}{\sqcup} \langle y_1, y_2, \dots, y_n \rangle = \langle x_1 \overset{D}{\sqcup} y_1, x_2 \overset{D^n}{\sqcup} y_2, \dots, x_n \overset{D^n}{\sqcup} y_n \rangle. \tag{2.4}$$

From here sets D^n, $n \leqslant |D|$ are CPOS as they contain the least element $\bot_{D^n}$ and any subset $X \subseteq D^n$ has least upper bound $\overset{D^n}{\sqcup} X \in D^n$. The obtained set $r(D)$ in (2.1) is CPOS too, while the union is maid by coupling (pasting) the least elements

$$\bot_{D^n}, \quad n = 1, 2, \dots, |D|.$$

From the conditions (2.3, 2.4) it follows that the operator r is monotonous, i.e., the directivity of the subset $X \subseteq D^n$, $n = 1, 2, \dots, |D|$ is not lost while extending the domain of definition from D to $r(D)$.

In addition from the equation

$$r(\overset{D}{\sqcup} X) = \overset{r(D)}{\sqcup} r(X) \tag{2.5}$$

it follows that the operator $r(D)$ is continuous.

Let us consider the second operator $p: D \to p(D)$ constructing a set of continuous functions:

$$p(D) = [D \to D]. \tag{2.6}$$

The ordering $\sqsubseteq_{p(D)}$ on $p(D)$ is stimulated by the ordering $\sqsubseteq_D$ on D :

$$\left.\begin{array}{c} x_1 \to y_1 \sqsubseteq_{p(D)} x_2 \to y_2 \\ \text{only when} \quad x_1 \sqsubseteq_D x_2 \quad \text{and} \quad y_1 \sqsubseteq_D y_2 \end{array}\right\}. \tag{2.7}$$

The set $p(D)$ is CPOS, because the least element $\perp_{p(D)} = (\perp_D \to \perp_D)$ is artificially introduced for it and any subset $X \subseteq p(D)$ has a least upper bound (supremum) $\overset{p(D)}{\sqcup} X \in [D \to D]$. The upper bound for the two elements $(x_1 \to y_1) \in p(D)$ and $(x_2 \to y_2) \in p(D)$ is calculated in the following way:

$$(x_1 \to y_1) \overset{p(D)}{\sqcup} (x_2 \to y_2) = (x_1 \overset{D}{\sqcup} x_2) \to (y_1 \overset{D}{\sqcup} y_2). \tag{2.8}$$

From the conditions (2.7, 2.8) it follows that the operator p remains monotonous, i.e., the directivity of the subsets $X \subseteq p(D)$ is not lost in transition from D to $p(D)$. In addition from the equation

$$p(\overset{D}{\sqcup} X) = \overset{p(D)}{\sqcup} p(X) \tag{2.9}$$

it follows that the operator p is continuous.

As the sets $r(D)$ and $p(D)$ are CPOS it is correct to expand the domain D of definition for the operators r and p in the following way:

$$r: \begin{cases} r(D) \to r(D) \\ p(D) \to r(D) \end{cases}, \qquad p: \begin{cases} r(D) \to p(D) \\ p(D) \to p(D) \end{cases}. \tag{2.10}$$

A composition of the monotonous and continuous functions is the monotonous continuous function, too. The domains of values of the operators r and p on (2.10) remain CPOS and are determined as:

$$\left.\begin{array}{l} r(D) = r(D) + r \circ p(r(D)) + \dots \\ p(D) = p(D) + p \circ r(p(D)) + \dots \end{array}\right\}. \tag{2.11}$$

As a result of the multiple use of the operators r and p the boundaries R and P are obtained as fixed point of continuous monotonous mapping of complete partially ordered sets in themselves. The fixed point theorem for a complete lattice has been proved by A. Tarski (Tarski, 1981). A variant of the fixed point theorem for CPOS is presented in (Barendregt, 1985). The boundared sets $\mathbf{R}$ and $\mathbf{P}$ are recursively determined by the operators r and p on the initial CPOS D of entities names. On the boundaries $\mathbf{R}$ and $\mathbf{P}$ the operators turn into operations $\overset{\circ}{r}$ and $\overset{\circ}{p}$ and the formulae (2.11) become:

$$\left.\begin{array}{l}\mathbf{R} = \overset{\circ}{r}\,(\mathbf{R}+\mathbf{P}) \\ \mathbf{P} = \overset{\circ}{p}\,(\mathbf{R}+\mathbf{P})\end{array}\right\}. \tag{2.12}$$

By disjunctive joining of $\mathbf{R}$ and $\mathbf{P}$ as CPOS the largest type $\mathbf{Q}$ is obtained:

$$\mathbf{Q} = \mathbf{R} + \mathbf{P}. \tag{2.13}$$

The type is CPOS and contains the possible types $\mathbf{R}$ and $\mathbf{P}$ constructed by means of operations $\overset{\circ}{r}$ and $\overset{\circ}{p}$ (2.12).

3. Conceptual and relational schemes of entities. With typed sets $\mathbf{Q}, \mathbf{R}$ and $\mathbf{P}$ at our disposal, we can determine subtype $Q \subseteq \mathbf{Q}$. We name it entity or object. Object Q is a new type and it can be defined by Q_i – commutation diagrams (Fig. 3.1).

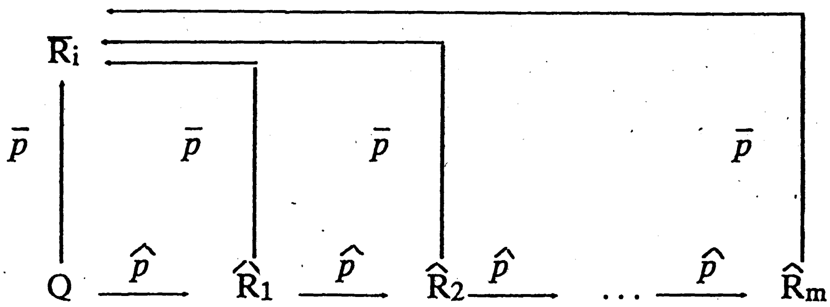

Fig. 3.1. Commutation Q_i – diagram.

Mappings in the diagram

$$\overline{p}: \begin{cases} Q \to \overline{R}_i, & \overline{R}_i \subseteq \mathbf{R}; \quad \overline{R}_i \in Q.\overline{\mathbf{R}}; \quad i = 1,2,\ldots,n \\ \widehat{R}_j \to \overline{R}_i, & \widehat{R}_j \subseteq \mathbf{R}; \quad \widehat{R}_j \in Q.\widehat{\mathbf{R}}; \quad j = 1,2,\ldots,m; \ m \leqslant n \end{cases} \tag{3.1}$$

are surjective and make up set $Q.\overline{\mathbf{P}} \subseteq \mathbf{P}$, wherein

$\overline{R}_i$, $i = 1,2,\ldots,n$ – domain of surjections $\overline{p}$ values, and we call it object Q internal relation;

$Q.\overline{\mathbf{R}}$ – set of object Q internal relationships;

$\widehat{R}_j$, $j = 1,2,\ldots,m$ – domain of surjections $\overline{p}$ definition, and we call them object Q identification relationship;

$Q.\widehat{\mathbf{R}}$ – set of object Q identification relationships.

Mappings in the diagram

$$p: \ Q \approx \widehat{R}_j; \quad \widehat{R}_j \subseteq \widehat{\mathbf{R}}; \quad \widehat{R}_j \in Q.\widehat{\mathbf{R}} \subseteq Q.\overline{\mathbf{R}}; \quad j = 1,2,\ldots,m \tag{3.2}$$

are bijective (as well as surjective and injective), make up set $Q.\mathbf{P} \subseteq \mathbf{P}$. If internal relation $R \in Q.\overline{\mathbf{R}}$ is unary, we name it "property" and denote: $T \subseteq D$. We use notation $Q.\mathbf{T}$ for the set of object Q unary relationships or properties. In order to indicate the context of relationship $R \in Q.\overline{\mathbf{R}}$, R may be denoted $Q.R$.

Though every identification relationship $R_j \in Q.\widehat{\mathbf{R}}$ is internal relationship $\overline{R}_j \in Q.\overline{\mathbf{R}}$, the reverse assertion can be false. The informational object is characteristic of its capacity, which is a set consisting of instance objects. Every instance object $q \in Q$ is identified through some instance relationship $\widehat{p}(\mathbf{q}) \in \widehat{R}_j$. We state that the conceptual scheme is defined for object $Q \in QS$ if the names from entity name universum D are assigned to Q and to internal relationships (components) of Q, i.e., if the following mapping is specified:

$$d: \begin{cases} QS \to D, \\ \{Q.R \mid Q \in QS; \quad Q.R \in Q.\overline{\mathbf{R}}\} \to D. \end{cases} \tag{3.3}$$

Specifying the object conceptual schemes of the problem domain in Q_i – commutation diagrams, we specify $Q \subseteq \mathbf{Q}$, $Q.\overline{R}_i \subseteq \mathbf{R}$ and $Q.\widehat{R}_j \subseteq \mathbf{R}$ which are partially ordered sets and are identified by their individualnames from D. The following equivalent relationships hold for object Q and names of its components and supremums:

$$\left.\begin{array}{l} d(Q) \equiv d(Q.\widehat{R}_j) \equiv \overset{D^l}{\sqcup}\ Q.\widehat{R}_j;\ Q.\widehat{R}_j \subseteq D^1,\ l \leqslant |D| \\ d(Q),\quad d(Q.\widehat{R}_j) \in D,\quad D^l = \underbrace{[D \times D \times \ldots \times D]}_{l} \subseteq r(D) \end{array}\right\}, \tag{3.4}$$

$$\left.\begin{array}{ll} d(Q.R_i) \equiv \overset{D^t}{\sqcup}\ Q.R_i; & Q.R_i \subseteq D^t,\ t \leqslant |D| \\ d(Q.R_i) \in D; & D^t = \underbrace{[D \times D \times \ldots \times D]}_{t} \subseteq r(D) \end{array}\right\}. \tag{3.5}$$

If object $Q \in QS$ from the problem area and their components $Q.R$ are bounded together by external inclusion relations E, we obtain the conceptual scheme of the information system:

$$\begin{aligned} M &= \{d(QS),\ d(\overline{\mathbf{R}}),\ d(\widehat{\mathbf{R}}),\ E\}, \qquad (3.6)\\ \overline{\mathbf{R}} &= \{Q.\overline{\mathbf{R}} | Q \in QS\}, \\ \widehat{\mathbf{R}} &= \{Q.\widehat{\mathbf{R}} | Q \in QS\}. \end{aligned}$$

Let us take IS conceptual scheme M with entities set QS in it, data base relational scheme $\mathbf{G}$, relation scheme $G \in \mathbf{G}$ and its attribute subsets $X \subseteq G$, which have mapping h so that:

$$h_s : QS \rightarrow \mathbf{G}, \tag{3.7}$$

$$h_Q : \begin{cases} Q.\overline{\mathbf{R}} \rightarrow G.X; \\ h_Q(Q.\widehat{R}_j) = G.X_j,\ j = 1, 2, \ldots, n; \\ h_Q(Q.\overline{R}_0) = G.X_0. \end{cases} \tag{3.8}$$

There $Q.\overline{R}_0$ – object Q internal relationship between all its properties which have been aggregated and involved into at least one identification relationship $Q.R_j$.

In this way, h_S and h_Q being specified, the functional structure of informational object $Q \in QS$ is represented on relational scheme $G \in \mathbf{G}$ by the set of functional dependences with equivalent left sides, i.e., by compound functional dependence $(X_1, X_2, \ldots, X_m) \rightarrow X_j$ (Maier, 1983).

4. Conclusions. Reflexiveness is an important feature of the given typing approach, and from this it is possible to assign the types by means of design operators r and p to the initial CPOS D without any initial types and pre-typing.

The transition from the initial set D to bounded sets $\mathbf{R}$ and $\mathbf{P}$ enables a set of entities names to be considered as structural types with respect to isomorphism: any structural entity in either a list or a function (2.12) and identified reflexively and cross-recursively (2.11) by means of monotonous and continuous operators r and p.

The given Q_i – commutation diagrams specify subtypes $Q \subseteq \mathbf{Q}$ with restricted list and functional structures, which are characterized by partially ordered sets $Q.\overline{\mathbf{R}}$ and $Q.\overline{\mathbf{P}}$, $Q.\widehat{\mathbf{P}}$. IS conceptual scheme is madded giving non-structural names to objects and to their components $Q.R$, and specifying external inclusion relations at the level of the sets. $(\mathbf{Q}, \sqsubseteq)$ is CPOS, we are able to define open set conditions for sets $Q \subseteq \mathbf{Q}$. Thus, Scoots topology can be introduced on basic set $\mathbf{Q}$, and classical topology operations can be used for manipulation in conceptual schemes.

Object $Q \subseteq QS$ are represented by relational schemes $G \in \mathbf{G}$ according to rules (3.7) and (3.8). These rules permit the functional structure of the conceptual scheme to be interpreted as compound functional dependence between corresponding attributes of subsets $X \subseteq G$. In this way , the tasks of functional analysis of data base relational scheme properties can be transferred to IS conceptual scheme level.

REFERENCES

Barendregt, H. (1985). In A.Kuzitchev (Ed.), *The Lamda Calculus. Its Syntax and Semantics.* Mir, Moscow (in Russian).

Chen, P.P-S. (1976). The entity-relationship model-toward a unified view of data. *ACM TODS,* 1(1), 9–36.

Daniel, M. (1990). Expert systems for database design support. *Proc. of the 13-th International Seminar on Database Management Systems.* Mamaia. 46–57.

Jasiukevičius A. (1986). The language of associative schemes for computer integrated management of complex objects. Material of seminar *Problems of*

Programming, Vilnius. 51–69 (in Russian).

Maier, D. (1983). *The Theory of Relational Database.* Computer Science Press, Rockville.

Tarski, A. (1981). A lattice-theoretical fixed point theorem and its syntax and semantics. *Studies in Logic and the Foundation of Mathematics*, **103**, North-Holland.

Received December 1993

B. Paradauskas received his degree of Candidate of Technical Sciences from Kaunas Politechnical Institute in 1974. He is the head of Information Systems department of Kaunas University of Technology. His research interests include conceptual modelling of data bases.