Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

The attached copy is furnished for non-commercial research and education use only.
Authors are permitted to post this version of the article to their personal websites or institutional repositories and to share with other researchers in the form of electronic reprints.
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to third party websites are prohibited.

For further information on
Serdica Bulgaricae Mathematicae Publicationes
and its new series Serdica Mathematical Journal
visit the website of the journal http://www.math.bas.bg/~serdica
or contact: Editorial Office
Serdica Mathematical Journal
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: serdica@math.bas.bg

SERDICA — Bulgaricae mathematicae publicationes
19 (1993) 121-133

Mathematical Structures - Informatics - Mathematical Modelling
Papers dedicated to Academician L. Iliev on the Occasion of his Eightieth Birthday

EQUALITY SCHEMES AND EQUALITY SCHEME DEPENDENCIES IN RELATIONAL DATABASES[†]

J. DEMETROVICS, L. RÓNYAI, HUA NAM SON

Dedicated to Academician Ljubomir Iliev
on the Occasion of his Eightieth Birthday

ABSTRACT. To each relation we associate an equality scheme (for short, ES) which is equivelent to the system of partitions (for short, SP) defined in [4]. The main operations of relational algebra are defined for ESs. By using the ESs we introduce a class of dependencies, called equality scheme dependencies (for short, ESDs). A set of TSDs is equivelent to a set of tuple generating dependencies (for short TGDs). A chase procedure is proposed for ESDs and for total 1-ESDs. We show also that in fact the approach is valid for the general case when we consider arbitrary, fixed standard realtions of data.

1. Introduction. For each database over an universe of data with predifined standard relationships we would like to extract out a scheme which represents well the given database.

At first, as an instance, let us consider the case of equality. The fact that a large portion of dependencies concerns the equalities of data shows that the equalities play an important role in the study of relational databases. Equality sets are used to investigate some types of dependencies [1, 4, 5, 6, 7], but are not powerful enough to represent even those dependencies whose definition needs only the equalities of data (for example, multivalued dependencies). In [6, 7] equality schemes (ESs, for short), were introduced which appear to be a more convenient tool than equality sets. We show that there is an 1-1 correspondence between the systems of partitions [4] and the equality schemes. This implies that all results concerning the equality of data in fact hold for a more general case when the equivalence of data is concerned. We define the main operation of relational algebra for ESs. By using ESs we can study several types of dependencies, such as multivalued dependencies [8], partition dependencies [4], etc. In this paper we introduce a clase of dependencies, called equality scheme dependencies

[†]The research was supported by Hungarian Foundation for Scientific Research Grant OTKA 2575.

(for short, ESDs) and prove that ESDs are TGDs mixed with EGDs. TGDs and EGDs which up to now have been studied separetely [2, 9] now can be studied in a unified setting. The chase proposed in [2] for total TGDs is modified for total 1-ESDs. A chase is also proposed for ESDs.

We show that this approach is valid for the general case when we consider arbitrary, fixed standard relation of data. Another instance is the case when the universe of data is partially ordered. An attempt to classify the databases based on the structure of the universe of data is demonstrated.

The paper is structured as follows: Sections 2-6 cover the case of equality. In Section 2 we give basic definitions of ESs, SPs and show the relationship between ESs and SPs. In section 3 we study the minimal ESs. In section 4 we define the main operations of relational algebra on the set of ESs. In Section 5 ESDs are introduced and studied. In Section 6 a chase for ESDs is given. Based on the chase proposed in [2] for total TGDs we present also a chase for total 1-ESDs. In Section 7 we deal with the general case. In 7.1. we demonstrate how this approach works for the case of inequality. In 7.2 the general case is studied. In 7.3 we demonstrate an attempt to classify the databases using the structural properties of the universes of data.

2. Equality schemes and systems of partitions. Let $\mathcal{U}$ be a finite set of attributes. The domain of $a \in \mathcal{U}$ is $Dom(a)$. A tuple t over $\mathcal{U}$ is a mapping $\mathcal{U} \to \cup_{a\in\mathcal{U}} Dom(a)$ such that $t[a] \in Dom(a)$ for all $a \in \mathcal{U}$. A relation over $\mathcal{U}$ is a set of tuples over $\mathcal{U}$. For a relation r let $Val(r,a) = \{t[a] | t \in r\}$ and $Val(r) = \cup_{a\in\mathcal{U}} Val(r,a)$.

Definition 1. *Let T be a set.*

1. *An equality scheme (for short,* ES*) of T is a couple $e = < T, l >$, where l is a mapping $T \times T \to 2^{\mathcal{U}}$ such that:*
   (i) (Symmetry) For all $t_1, t_2 \in T$: $l(t_1,t_2) = l(t_2,t_1)$,
   (ii) (Triangle condition) $\forall t_1, t_2, t_3 \in T$: $l(t_1,t_2) \cap l(t_2,t_3) \subseteq l(t_1,t_3)$,
   (iii) $l(t_1,t_2) \neq \mathcal{U}$ for all $t_1,t_2 \in T$, $t_1 \neq t_2$ and $l(t,t) = \mathcal{U}$ for all $t \in T$.
   The ESs *are denoted by $e, f, \ldots$. An* ES *of T is finite if T if finite.*

2. *For a relation $r = \{t_1, \ldots, t_n\}$ over $\mathcal{U}$ put $e_r = < r, l >$ where $l(t_i,t_j) = \{a \in \mathcal{U} | t_i[a] = t_j[a]\}$. One can verify that e_r is an* ES. *We call e_r the* ES *of r.*

3. *Let $e_i = < T_i, l_i >$ be* ESs*, $i = 1,2$, and h be a mapping $T_1 \to T_2$. We write $h : e_1 \xrightarrow{*} e_2$ if $l_1(t_1,t_2) \subseteq l_2(h(t_1),h(t_2))$ for all $t_1,t_2 \in T_1$. If $h(t_1) = h(t_2)$ for all $t_1,t_2 \in T_1$ then we say that h is trivial. If $h : e_1 \xrightarrow{*} e_2$ is not trivial then we write also $e_1 \leq_h e_2$. We write $e_1 \leq_* e_2$ if $e_1 \leq_h e_2$ for some h. We write $e_1 \equiv_* e_2$ if $e_1 \leq_* e_2$ and $e_2 \leq_* e_1$.*

4. *We write $e_1 = e_2$ iff there is a* 1-1, *onto mapping $h : T_1 \to T_2$ such that $l_1(t_1,t_2) = l_2(h(t_1),h(t_2))$ for all $t_1,t_2 \in T_1$.*

The elements of T are referred to as the names of tuples in databases.

Proposition 1. *Let e be an* ES *and r_2, r_2 be relations. Then:*

1. *There exits a relation r such that $e = e_r$.*

2. *$e_{r_1} = e_{r_2}$ iff there are* 1-1 *onto mapping $i_a : Val(r_1, a) \to Val(r_2, a)$ such that $t_2 \in r_2 \iff \exists t_1(\forall a \in \mathcal{U} : t_2[a] = i_a(t_1[a]))$.*

The proof of (1) for the general case is the same as in the case where e is finite (see[6]). We omit the simple proof of (2).

Definition 2. *Let T be a set and $\mathcal{U}$ be a finite set of attributes.*

1. *A system of partitions* (SP, *for short*) *of T over $\mathcal{U}$ is a system $\mathcal{P} = \{P_a | a \in \mathcal{U}\}$, where P_a is a partition of T, $a \in \mathcal{U}$ and $\Pi_{a \in \mathcal{U}} P_a = \mathcal{P}_0$.*

2. *Let $r = \{t_1, \ldots, t_n\}$ be a relation over $\mathcal{U}$. For $a \in \mathcal{U}$ let P_a be the partition of r that is defined by the equivalence: $t_i \equiv_{P_a} t_j \iff t_i[a] = t_j[a]$. Put $\mathcal{P}_r = \{P_a | a \in \mathcal{U}\}$. $\mathcal{P}_r$ is called the system of partitions of r.*

3. *Let $\mathcal{P}_i = \{P_x^i | x \in \mathcal{U}\}$ be an* SP *of T_i, $i = 1,2$ and h be a non-trivial mapping $T_1 \to T_2$. We write $h : \mathcal{P}_1 \overset{*}{\to} \mathcal{P}_2$ iff $t \equiv_{P_a^1} t' \implies h(t) \equiv_{P_a^2} h(t')$ for all $a \in \mathcal{U}$, $t, t' \in T$. In this case we write also $\mathcal{P}_1 \leq_h \mathcal{P}_2$. We write $\mathcal{P}_1 \leq_* \mathcal{P}_2$ iff $\mathcal{P}_1 \leq_h \mathcal{P}_2$ for some $h : T_1 \to T_2$. $\mathcal{P}_1 \equiv_* P_2$ stand for $\mathcal{P}_1 \leq_* P_2$ and $\mathcal{P}_2 \leq_* P_1$. We write $\mathcal{P}_1 = \mathcal{P}_2$ iff there is a* 1-1, *onto mapping $h : T_1 \to T_2$ such that $t \equiv_{P_a^1} t' \iff h(t) \equiv_{P_a^2} h(t')$ for all $a \in \mathcal{U}$, $t, t' \in T_1$.*

The SPs defined as in (1), (2) have been proposed first in [4]. We have:

Theorem 1. *There exists a* 1-1, *onto mapping F from the set of* ESs *to the set of* SPs *over $\mathcal{U}$, such that:*

(i) $F(e_r) = \mathcal{P}_r$ for all relation r.

(ii) $e_1 \leq_* e_2$ iff $F(e_1) \leq_* F(e_2)$.

P r o o f. Let $e =< T, l >$ be an ES. Put $F(e) = \{P_a | a \in \mathcal{U}\}$ where $t_1 \equiv_{P_a} t_2 \iff a \in l(t_1, t_2)$ for all $t_1, t_2 \in T$. It is easy to see that $F(e)$ is an SP. F is a 1-1, onto mapping that satisfies (i), (ii). □

Since SPs represent the equivalences of data, ESs in fact characterize the equivalences of the data in databases.

3. Minimal ESs.

Definition 3. *Let $e_i =< T_i, l_i >$ be ES over $\mathcal{U}$, $i = 1,2$. We write $e_1 \subseteq e_2$ if $T_1 \subseteq T_2$ and l_1 is the restriction of l_2 on $T_1 \times T_1$. We say that e_2 is a minimal* ES *iff there is no non-trivial e_1 such that $e_1 \subset e_2$ and $e_1 \equiv_* e_2$. e_1 is a minimal* ES *of e_2 iff $e_1 \subseteq e_2$, $e_1 \equiv_* e_2$ and e_1 is minimal.*

The following Theorem 2,3 characterize the minimal ESs:

Theorem 2. *An* ES*s is minimal iff all non-trivial maps $h : e \overset{*}{\to} e$ are surjective.*

We omit the simple proof. For finite ESs we have:

Theorem 3.

1. *A finite* $e =< T, l >$ *is minimal* ES *then all non-trivial* $h : e \xrightarrow{*} e$ *are* 1-1 *mappings. In this case for each* $t \in T$ *there is an integer* m *such that* $h^{(m)}(t) = t$ *where* $h^{(1)}(t) = h(t)$, $h^{(n+1)}(t) = h^{(n)}(h(t))$.

2. *Let* e_1, e_2 *be finite, minimal* ESs. *Then* $e_1 \equiv_* e_2$ *iff* $e_1 = e_2$.

Proof. 1. is obvious, because h is a permutation of the finite set T.

2. Let $e_i =< T_i, l_i >$ where T_i is finite, $i = 1, 2$. It is evident that $e_1 = e_2$ implies $e_1 \equiv_* e_2$. Conversly, suppose that $e_1 \equiv_* e_2$ where $h : e_1 \xrightarrow{*} e_2$ and $k : e_2 \xrightarrow{*} e_1$. We have $k \circ h : e_1 \xrightarrow{*} e_2$ and $h \circ k : e_2 \xrightarrow{*} e_1$. By Theorem 2 and (1) of this theorem $h \circ k$ and $k \circ h$ are 1-1, onto mappings, i.e. h, k are 1-1, onto mappings. Moreover, for $t_1, t_2 \in T_1$ there are m_1, m_2 such that $(k \circ h)^{(m_1)}(t_1) = t_1$, $(k \circ h)^{(m_2)}(t_2) = t_2$. Put $m = m_1 m_2$. We have $l_1(t_1, t_2) \subseteq l_2(h(t_1), h(t_2))$ and $l_1(t_1, t_2) = l_1((k \circ h)^{(m)}(t_1), (k \circ h)^{(m)}(t_2)) \supseteq l_2(h(t_1), h(t_2))$, i.e. $e_1 = e_2$. □

4. The operations of the relational algebra on ESs. We define the main operations of the relational algebra for ESs. We use the notations of [12].

1. *Union.* The ESs $e_1 =< T_1, l_1 >$, $e_2 =< T_2, l_2 >$ are *unifiable* if $l_1(t_1, t_2) = l_2(t_1, t_2)$ for all $t_1, t_2 \in T_1 \cap T_2$. In this case let us put $T = T_1 \cup T_2$ and

$$l(t_1, t_2) = \begin{cases} \bigcup\limits_{t \in T_1 \cap T_2} (l_1(t_1, t) \cap l_2(t, t_2)), & \text{if } T_1 \cap T_2 \neq \emptyset; \\ \emptyset, & \text{otherwise.} \end{cases}$$

e is an ES. We call e the *union* of e_1, e_2 and denote by $e_1 \cup e_2$. The *union* of $\{e_i | i \in I\}$ is defined by induction and is denoted by $\bigcup\limits_{i \in I} e_i$.

2. *Set difference.* Let $e_1 =< T_1, l_1 >$, $e_2 =< T_2, l_2 >$ be unifiable ESs. The *difference* of e_1, e_2, denoted by $e_1 - e_2$, is $e =< T_1 \setminus T_2, l >$ where l is the restriction of l_1 into $T_1 \setminus T_2$.

3. *Cartesian product of type* 1. Let $e_1 =< T_1, l_1 >$, $e_2 =< T_2, l_2 >$ be ESs over $\mathcal{U}_1, \mathcal{U}_2$, respectively, $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$. The *Cartesian product of type* 1 of e_1, e_2, denoted by $e_1 \times e_2$, is $e =< T_1 \times T_2, l >$ where for all (t_1, t_2), $(t_1', t_2') \in T_1 \times T_2$:

$$l((t_1, t_2), (t_1', t_2')) = l_1(t_1, t_1') \cup l_2(t_2, t_2').$$

e is an ES. The Cartesian product of type 1 of two relations r_1, r_2 denoted by $r_1 \times r_2$, is defined as the Cartesian product of r_1, r_2 in [12].

4. *Cartesian product of type* 2. Let $e_1 =< T_1, l_1 >$, $e_2 =< T_2, l_2 >$ be ESs. The *Cartesian product of type* 2 of e_1, e_2, denoted by $e_1 \otimes e_2$, is $e =< T_1 \times T_2, l >$ where for all (t_1, t_2), $(t'_1, t'_2) \in T_1 \times T_2$:

$$l((t_1, t_2), (t'_1, t'_2)) = l_1(t_1, t'_1) \cap l_2(t_2, t'_2).$$

e is an ES. For $a \in \mathcal{U}$ let a' be a new attribute such that $Dom(a') = Dom(a) \times Dom(a)$. For two relations $r_1 = \{t^1_1, \ldots, t^1_n\}$, $r_2 = \{t^2_1, \ldots, t^2_m\}$ over $\mathcal{U}$ the Cartesian product of type 2 of r_1, r_2, denoted by $r_1 \otimes r_2$, is the relation $r = \{(t^1_i, t^2_j) | 1 \leq i \leq n,\ 1 \leq j \leq m\}$, where $(t^1_i, t^2_j)[a'] = (t^1_i[a], t^2_j[a])$.

5. *Selection.* Let $e =< T, l >$ be an ES and $T_1 \subseteq T$. The *selection* of e by T_1, denoted by $\psi_{T_1}(e)$, is $e_1 =< T_1, l_1 >$ where l_1 is the restriction of l into $T_1 \times T_1$.

6. *Projection.* Let $e =< T, l >$ be an ES and $X \subseteq \mathcal{U}$, $X \neq \emptyset$. We write $t_1 \equiv t_2$ if $l(t_1, t_2) \supseteq X$. $\equiv$ is an equivalence on T. Let $[t]$ denote the equivalent class of t. Put $T_1 = \{[t] | t \in T\}$ and $t_1 : T_1 \times T_1 \to 2^{\mathcal{U}}$ where:

$$l_1([t_1], [t_2]) = \bigcup_{t'_1 \in [t_1], t'_2 \in [t_2]} l(t'_1, t'_2).$$

$e_1 =< T_1, l_1 >$ is an ES. e_1 is the *projection* of e into X and is denoted by $\pi_X(e)$.

7. *Join.* We define the join of two ESs. Join of more ESs can be defined by induction. For $i = 1, 2$ let $e_i =< T_i, l_i >$ be an ES over $\mathcal{U}_i$, respectively. For (t_1, t_2), $(t'_1, t'_2) \in T_1 \times T_2$ put

$$\begin{aligned} m_1(t_1, t'_1) &= l_1(t_1, t'_1) \cap (\mathcal{U}_1 \setminus \mathcal{U}_2), \\ m_2((t_1, t_2), (t'_1, t'_2)) &= l_1(t_1, t'_1) \cap l_2(t_2, t'_2), \\ m_3(t_2, t'_2) &= l_2(t_2, t'_2) \cap (\mathcal{U}_2 \setminus \mathcal{U}_1), \\ l((t_1, t_2), (t'_1, t'_2)) &= m_1(t_1, t'_1) \cup m_2((t_1, t_2), (t'_1, t'_2)) \cup m_3(t_2, t'_2). \end{aligned}$$

Lemma 1. *Let $T = T_1 \times T_2$ and let l be defined as above. Then $e =< T, l >$ is an ES on $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2$.*

The routine proof is omitted. We call e the *join* of e_1, e_2 and denote $e = J(e_1, e_2)$. Let $\rho \subseteq \Pi^m_{i=1} T_i$. Then $J_{[\rho]}(e_1, e_2) = \psi_\rho(J(e_1, e_2))$ is called the *join under the condition* ρ of e_1, e_2.

Proposition 2. *Let e_i be unifiable* ES*s for $i \in I$. Then*

1. $e_i \leq_* \bigcup_{j \in I} e_j$ *for all* $i \in I$.
2. *If e_i are* ES*s of disjoint sets and $e_i \leq_* e$ for all $i \in I$ then* $\bigcup_{i \in I} e_i \leq_* e$.

We omit the simple proof. We have:

Theorem 4. *Let r_1, r_2 be relatios over $\mathcal{U}$. Then:*

1. e_{r_1}, e_{r_2} *are unifiable* ESs *and* $e_{r_1} \cup e_{r_2} \leq_* e_{r_1 \cup r_2}$.
2. $e_{r_1} - e_{r_2} = e_{r_1 - r_2}$.
3. $e_{r_1} \times e_{r_2} = e_{r_1 \times r_2}$.
4. $e_{r_1} \otimes e_{r_2} = e_{r_1 \otimes r_2}$.
5. *If* $r_1 \subseteq r_2$ *then* $\psi_{r_1}(e_{r_2}) = e_{r_1}$.
6. *For* $X \subseteq \mathcal{U}$, $X \neq \emptyset$ *we have* $\pi_X(e_{r_1}) = e_{\pi_X(r_1)}$.
7. *Let* $r_i = \{t^i_j | j = 1, \ldots, n_i\}$ *be relation over* $\mathcal{U}_i$, $i = 1, 2$, *respectively, and* ρ *be the "natural condition" on* $r_1 \times r_2$*:*

$$\rho = \{(t_1, t_2) | t_1 \in r_1, t_2 \in r_2 : t_1[\mathcal{U}_1 \cap \mathcal{U}_2] = t_2[\mathcal{U}_1 \cap \mathcal{U}_2]\}.$$

We have $J_{[\rho]}(e_{r_1}, e_{r_2}) = e_{r_1 \bowtie r_2}$.

We omit the routine proofs.

Example 1. Let $r_i, i = 1, 2, 3$ be relations in Fig. 1, 2, 3. Then $\mathcal{P}_{r_1}$, $\mathcal{P}_{r_2}$ and $\mathcal{P}_{r_1 \cup r_2} = \mathcal{P}_{r_1} \cup \mathcal{P}_{r_2}$ are described in Fig. 4, 5, 6, respectively. e_{r_1} and $J[\rho](e_{r_1}, e_{r_3}) = e_{r_1 \bowtie r_3}$ are described in Fig. 7, 8, respectively.

	a	b	c	d
t_1	a_0	b_0	c_0	d_0
t_2	a_0	b_1	c_0	d_0
t_3	a_1	b_0	c_0	d_1
t_4	a_1	b_0	c_1	d_0

Fig. 1. Relation r_1.

	a	b	c	d
v_1	a_2	b_2	c_2	d_2
v_2	a_2	b_3	c_3	d_2
v_3	a_4	b_2	c_3	d_4

Fig. 2. Relation r_2.

	c	d	e
w_1	c_0	d_0	e_0
w_2	c_0	d_1	e_0
w_3	c_0	d_1	e_1

Fig. 3. Relation r_3.

$a: [t_1, t_2], [t_3, t_4];$
$b: [t_1, t_3, t_4], [t_2];$
$c: [t_1, t_2, t_3], [t_4];$
$d: [t_1, t_2, t_4], [t_3];$

Fig. 4. $\mathcal{P}_{r_1}$.

$a: [v_1, v_2], [v_3];$
$b: [v_1, v_3], [v_2];$
$c: [v_2, v_3], [v_1];$
$d: [v_1, v_2], [v_3];$

Fig. 5. $\mathcal{P}_{r_2}$.

$a: [t_1, t_2], [t_3, t_4], [v_1, v_2], [v_3];$
$b: [t_1, t_3, t_4], [t_2], [v_1, v_3], [v_2];$
$c: [t_1, t_2, t_3], [t_4], [v_2, v_3], [v_1];$
$d: [t_1, t_2, t_4], [t_3], [v_1, v_2], [v_3];$

Fig. 6. $\mathcal{P}_{r_1 \cup r_2} = \mathcal{P}_{r_1} \cup \mathcal{P}_{r_2}$.

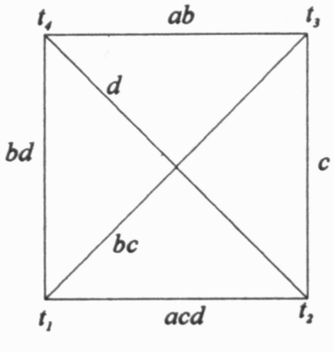

Fig. 7. e_{r_1}.

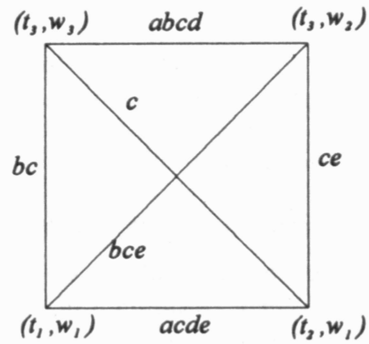

Fig. 8. $J_{[\rho]}(e_{r_1}, e_{r_3})$.

One can see that $h : \mathcal{P}_{r_2} \xrightarrow{*} \mathcal{P}_{r_1}$ where h is the mapping that takes v_i to t_i, $(i = 1, 2, 3)$. In other words we have $\mathcal{P}_{r_2} \leq_h \mathcal{P}_{r_1}$ and $e_{r_2} \leq_h e_{r_1}$. I (Fig. 9) constructed as in the proof of (1), Proposition 1, is a sample relation of e_{r_1}.

I	a	b	c	d_0
w_1	t_1t_2	$t_1t_3t_4$	$t_1t_2t_3$	$t_1t_2t_4$
w_2	t_2t_1	t_2	$t_2t_1t_3$	$t_2t_1t_4$
w_3	t_3t_4	$t_3t_1t_4$	$t_3t_1t_2$	t_3
w_4	t_4t_3	$t_4t_1t_3$	t_4	$t_4t_1t_2$

Fig. 9. A sample relation of e_{r_1}.

5. ES dependencies.

Here we assume that $Dom(a) \cap Dom(b) = \emptyset$ for $a \neq b$.

Definition 4. *An* ES *dependencies (*ESD*, for short) is an expression of the form* $e_1 \leq_h e_2$ *where* e_1, e_2 *are* ESs *and* $h : e_1 \xrightarrow{*} e_2$. *An* ESD $e_1 \leq_h e_2$ *is finite if* e_1, e_2 *are finite. A relation* r *satisfies* $\alpha = e_1 \leq_h e_2$ *iff for all* $k_1 : e_1 \xrightarrow{*} e_r$ *there exists* $k_2 : e_2 \xrightarrow{*} e_r$ *such that* $k_1 = k_2 \circ h$.

Let $\mathcal{C}_1, \mathcal{C}_2$ be sets of dependencies and α be a dependency. We write $\mathcal{C}_1 \vdash \alpha$ if r satisfies α for all r that satisfies $\mathcal{C}_1$. We write $\mathcal{C}_1 \vdash \mathcal{C}_2$ if $\mathcal{C}_1 \vdash \alpha$ for all $\alpha \in \mathcal{C}_2$. We write $\mathcal{C}_1 \equiv \mathcal{C}_2$ if $\mathcal{C}_1 \vdash \mathcal{C}_2$ and $\mathcal{C}_2 \vdash \mathcal{C}_1$.

The details of TGDs and EGDs can be found in [4, 9]. TGDs and EGDs generalize many important types of dependencies. To $a \in \mathcal{U}$ let $V(a)$ denote a set of variables whose domain is $Dom(a)$. Suppose that $V(a) \cap V(b) = \emptyset$ for $a \neq b$. For $X \subseteq \mathcal{U}$ put $V(X) = \Pi_{a \in X} V(a)$. A *tableau* over $\mathcal{U}$ is a pair $< w, I >$ where $w \in V(\mathcal{U})$, $I \subseteq V(\mathcal{U})$

and I is finite. A *valuation* of I is a mapping $\rho : \bigcup_{a\in\mathcal{U}} \{w[a] | w \in I\} \to \bigcup_{a\in\mathcal{U}} Dom(a)$ such that $\rho(w[a]) \in Dom(a)$. In the natural way ρ can be extended into a mapping $I \to Dom(\mathcal{U})$. A TGD is a pair $\alpha =< I_2, I_1 >$ where I_1, I_2 are finite subsets of $V(\mathcal{U})$, $I_1 \subseteq I_2$. A relation r satisfies α if for all valuations ρ_1 such that $\rho_1(I_1) \subseteq r$ there exists an extension ρ_2 of ρ_1 onto I_2 such that $\rho_2(I_2) \subseteq r$. An EGD is a pair $\beta =< x = y, I >$ where I is a finite subset of $V(\mathcal{U})$, $x, y \in V(a)$ for some $a \in \mathcal{U}$. A relation r satisfies β if for all valuations ρ such that $\rho(I) \subseteq r$ there must be $\rho(x) = \rho(y)$. We have:

Lemma 2. *Let r be a relation and $I \subseteq V(\mathcal{U})$.*

1. For a valuation ρ of I and $h_\rho : I \to Dom(\mathcal{U})$ where $h_p(t)[a] = \rho(t[a])$ we have $\rho(I) \subseteq r$ iff $h_p : e_I \overset{*}{\to} e_r$.

2. Let $h : e_I \overset{*}{\to} e_r$ and ρ_h be a valuation of I where $\rho_h(w[a]) = h(w)[a]$ for all $a \in \mathcal{U}$. Then $\rho_h(I) \subseteq r$.

We have:

Theorem 5. EGD*s and* TGD*s are finite* ESD*s.*

Proof. 1. Let $\beta =< x = y, I >$ be an EGD where $I = \{v_1, \ldots, v_n\}$. Let $J = \{w_1, \ldots, w_n\}$ where w_i is obtained from v_i by identifying x with y. Put $h : I \to J$ where $h(v_i) = w_i$. One can verify that $h : e_I \overset{*}{\to} e_J$. Put $\tau = e_I \leq_h e_J$. The proof that $\{b\} \equiv \{\tau\}$ is left for the readers.

2. Let $\alpha =< I_2, I_1 >$ be an TGD where $I_1 = \{v_1, \ldots, v_n\}$, $I_2 = \{v_1, \ldots, v_n, \ldots, v_m\}$. Put $\tau = e_{I_1} \leq_i e_{I_2}$ where i is the identity mapping $I_1 \to I_2$. One can verify that $\{\alpha\} \equiv \{\tau\}$. $\square$

Example 2. 1. Functional dependency $X \to Y$ can be represented by ESD $\alpha = e_1 \leq_h e_2$ where $e_i =< T_i, l_i >$, $T_i = \{u_i, v_i\}, l_1(u_1, v_1) = X, l_2(u_2, v_2) = XY$ and $h(u_1) = u_2, h(v_1) = v_2$.

2. Multivalued dependency $X \to\to Y$ is represented by ESD $\alpha = e_1 \leq_h e_2$ where $e_i =< T_i, l_i >$, $T_1 = \{u_1, v_1\}$, $T_2 = \{u_2, v_2, w_2\}$, $l_1(u_1, v_1) = X$, $l_2(u_2, v_2) = X$, $l_2(u_2, w_2) = XY$, $l_2(v_2, w_2) = X(\mathcal{U} \setminus XY)$ and $h(u_1) = u_2$, $h(v_1) = v_2$.

3. Join dependency $*[X_1, \ldots, X_m]$ is represented by ESD $\pi = e_1 \leq_h e_2$ where $e_i =< T_i, l_i >$, $T_1 = \{t_1, \ldots, t_m\}$, $T_2 = T_1 \cup \{t_0\}$, $l_1(t_i, t_j) = l_2(t_i, t_j) = X_i \cap X_j$ for $1 \leq i, j \leq m$, $l_2(t_0, t_j) = X_j$ and h is the identical mapping of T_1 into T_2.

We have:

Theorem 6. *For a set of finite* ESD*s C_1 we can find a set of* EGD*s and* TGD*s C_2 and vice versa, for a set of* EGD*s and* TGD*s C_2 we can find a set of finite* ESD*s C_1, such that $C_1 \equiv C_2$.*

Proof. Let C_1 be a set of finite ESDs. Without loosing of generality, suppose that C_1 consists of one ESD, $C_1 = \{\alpha\}$ where $\alpha = e_1 \leq_h e_2$. By Proposition 1 there there are $I_i \supseteq V(\mathcal{U})$ such that $e_i = e_I$, $i = 1, 2$. Suppose that $I_1 = \{t_1, \ldots, t_n\}$. For $a \in \mathcal{U}$, $i, j = 1, \ldots, n$ if $h(t_i) = h(t_j)$ then put

$$\begin{array}{ll}\alpha_{a,i,j} & =< t_i[a] = t_j[a], I_1 >, \\ \mathrm{EGD}(\alpha) & = \{\alpha_{a,i,j} | a \in \mathcal{U}, i, j : h(t_i) = h(t_j)\}.\end{array}$$

Moreover, put $e_3 = \psi_{h(I_1)}(e_2) =< I_3, l_3 >$, $\mathrm{TGD}(\alpha) = \{< I_2, I_3 >\}$ and $\mathcal{C}_2 = \mathrm{EGD}(\alpha) \cup \mathrm{TGD}(\alpha)$. We can verify that $\mathcal{C}_1 \equiv \mathcal{C}_2$. The second part of theorem follows from Theorem 5. □

6. Chase of ESDs and total 1-ESDs. Let $h_i : e_1^i \xrightarrow{*} e_2^i$ where e_j^i are ESs of pairwise disjoint sets T_j^i, $(i = 1, \ldots, n; j = 1, 2)$. For $h : \bigcup_{i=1}^{n} T_1^i \to \bigcup_{i=1}^{n} T_2^i$ where $h(t) = h_i(t)$ if $t \in T_1^i$ we have $h : \bigcup_{i=1}^{n} e_1^i \xrightarrow{*} \bigcup_{i=1}^{n} e_2^i$. Denote $h = \bigcup_{i=1}^{n} h_i$.

Definition 5. *Union of* α_i *is* ESD $\alpha = e_1 \leq_h e_2$ *where* $e_j = \bigcup_{i=1}^{n} e_j^i$ *for* $j = 1, 2$ *and* $h = \bigcup_{i=1}^{n} h_i$.

Remark that union is defined only for $\alpha_i = e_1^i \leq_{h_i} e_2^i$ where e_j^i are ESs of pairwise disjoint sets. A "renaming" can turn an arbitrary ESD into a ESD satisfying this condition.

Theorem 7. *Let* $\mathcal{C} = \{\alpha_i | i \in I\}$ *be a set of* ESD*s. Then* $\mathcal{C} \equiv \bigcup_{i \in I} \alpha_i$.

Proof. Suppose that $\alpha_i = e_1^i \leq_h e_2^i, e_j^i$ is an ES of T_j^i for $i \in I$, $j = 1, 2$. If r satisfies $\mathcal{C}$ and $k : \bigcup_{i \in I} e_1^i \xrightarrow{*} e_r$ then denote the restriction of k into e_1^i by k_i, we have $k_i : e_1^i \xrightarrow{*} e_r$. Since r satisfies $\mathcal{C}$ there must be $k_i' : e_2^i \xrightarrow{*} e_r$ such that $k_i = k_i' \circ h_i$. By Proposition 2 we have $k = \bigcup_{i \in I} k_i' : \bigcup_{i \in I} e_2^i \xrightarrow{*} e_r$ and $k = \bigcup_{i \in I} k_i' \circ \bigcup_{i \in I} h_i$, i.e. r satisfies $\bigcup_{i \in I} \alpha_i$.

Conversely, suppose that r satisfies $\bigcup_{i \in I} \alpha_i$. For a fixed i let $k_i : e_1^i \xrightarrow{*} e_r$ and $w_0 = k_i(t_0)$ where t_0 is a fixed element in T_1^i. Put $K_i : \bigcup_{i \in I} T_1^i \to \bigcup_{i \in I} T_2^i$ where

$$K_i(t) = \begin{cases} k_i(t), & \text{if } t \in T_1^i, \\ w_0, & \text{otherwise.} \end{cases}$$

We can verify that $K_i : \bigcup_{i \in I} e_1^i \xrightarrow{*} e_r$. There must be $K_i' : \bigcup_{i \in I} e_2^i \xrightarrow{*} e_r$ such that

$K_i = K'_i \circ \left(\bigcup_{i\in I} h_i\right)$. Let k'_i be the restriction of K'_i into T_2^i. We have $k'_i : e_2^i \xrightarrow{*} e_r$. For $t \in T_1^i$ we have $k'_i \circ h_i(t) = K'_i \circ \left(\bigcup_{j\in I} h_j\right)(t) = K_i(t) = k_i(t)$, i.e. r satisfies α_i. Since this holds for all $i \in I$, r satisfies $\mathcal{C}$. □

Next we define a chase for total ESDs. The approach is based on the method proposed in [2] for total TGDs.

Definition 6. *Let $\alpha = e_1 \leq_h e_2$ be an ESD where $e_i =< T_i, l_i >$, $i = 1, 2$, respectively. Then:*

1. *α is a 1-ESD if $|T_2 \setminus h(T_1)| = 1$.*
2. *α is a total if for all $t_2 \in T_2 \setminus h(T_1)$: $\bigcup_{t_1\in T_1} l_2(t_2, h(t_1)) = \mathcal{U}$.*

Lemma 3. *If $\alpha_i, i = 1, \ldots, n$, are total ESDs then $\bigcup_{i=1}^{n} \alpha_i$ is a total ESD.*

Lemma 4. *An ESD α is equivalent to a set of total 1-ESDs iff α is equivalent to some total ESD $\beta = e_1 \leq_h e_2$ where e_i are ESs of T_i, respectively, and $|T_2 \setminus h(T_1)| \geq 1$.*

Lemma 5. *Let $\alpha = e_1 \leq_h e_2$, $\beta = e_3 \leq_k e_4$ be 1-ESDs where $e_i =< T_i, l_i >$, $i = 1, \ldots, 4$, $T_2 = h(T_1) \cup \{t_0^1\}$, $T_4 = k(T_3) \cup \{t_0^2\}$. Then*

1. *Put $T_5 = T_3 \times T_1$, $l_5 : T_5 \times T_5 \to 2^{\mathcal{U}}$ where*

$$l_5((t_3,t_1),(u_3,u_1)) = \begin{cases} l_1(t_1,u_1), & \text{if } t_3 = u_3, \\ l_3(t_3,u_3) \cap l_1(t_1,u_1) \cap \\ \cap l_2(t_0^1, h(t_1)) \cap l_2(t_0^1, h(u_1)), & \text{otherwise.} \end{cases}$$

Then $e_5 =< T_5, l_5 >$ is an ES.

2. *Put $T_6 = k(T_3) \times h(T_1) \cup \{(t_0^2, t_0^1)\}$, where for $t_4, u_4 \in k(T_3), t_2, u_2 \in h(T_1)$,*

$$l_6((t_4,t_2),(u_4,u_2)) = \begin{cases} l_2(t_2,u_2), & \text{if } t_4 = u_4, \\ l_4(t_4,u_4) \cap l_2(t_2,u_2) \cap \\ \cap l_2(t_0^1, t_2) \cap l_2(t_0^1, u_2), & \text{otherwise.} \end{cases}$$

and $l_6((t_0^2, t_0^1),(u_4,u_2)) = l_4(t_0^2, u_4) \cap l_2(t_0^1, u_2)$. Then $e_6 =< T_6, l_6 >$ is an ES. Moreover, if we put $k' : T_5 \to T_6$ where $k'(t_3,t_1) = (k(t_3), h(t_1))$ then $k' : e_5 \xrightarrow{*} e_6$ and $e_5 \leq_{k'} e_6$ is a 1-ESD. Denote $e_5 \leq_{k'} e_6$ by $\beta \circ \alpha$.

3. *If α , β are total then $\beta \circ \alpha$ are total.*

We omit the routine, long proofs.

Definition 7. (Product of total 1-ESDs) *Let α, β be* 1-ESD*s. We call $\beta \circ \alpha$ the product of α, β.*

The set of total 1-ESDs with the product defined in Definition 7 forms a monoid.

Theorem 8. *Let $C = \{\alpha, \ldots, \beta\}$ be a set of total* 1-ESD*s. Then $C \equiv \alpha \circ \ldots \circ \beta$.*

The long proof will be presented in a full version of this paper.

7. Generalization. Theorem 1 shows that though ESs are defined by the equalities of data in databases they in fact are equivalent to SPs which are determined by the equivalences of data. Thus ESs in fact characterize the equivalence of data. This implies that the results known up to now for FDs, JDs, TGDs, EGDs, etc., hold for a larger class of dependencies. In the followings we show that our approach described in Sections 1-6 is valid for other predefined structures of data. In 7.1 we deal with inequalities, or with partial orderings of data. In 7.2 we consider the general case.

7.1 Schemes representing the inequalities of data.

Definition 8.

1. *An inequality scheme of T over $\mathcal{U}$ is a pair $i = < T, l >$ where $l : T \times T \to 2^{\mathcal{U}}$ that satisfies the followings:*

(i) $l(t_1, t_2) \cup l(t_2, t_1) = \mathcal{U}$ for all $t_1, t_2 \in T$, and

(ii) (Triangle condition) $\forall t_1, t_2, t_3 \in T : l(t_1, t_2) \cap l(t_2, t_3) \subseteq l(t_1, t_3)$.

2. *Let $r = \{t_1, t_2, \ldots, t_n\}$ be a relation over $\mathcal{U}$ where $Dom(a)$ is completely ordered by $\leq_a$ for all $a \in \mathcal{U}$. The inequality scheme of r is $i_r = < r, l_r >$ where $l_r(t_1, t_2) = \{a \in \mathcal{U} | t_1[a] \leq_a t_2[a]\}$. One can verify easily that i_r is an inequality scheme.*

7.2 Schemes representing an arbitrary standard relationships of data.

In this section for a set A let $A^{(n)} = \underbrace{(A \times \ldots \times A)}_{\text{n times}}$ and $A^{(*)} = \bigcup_{n=1}^{\infty} A^{(n)}$.

For $a \in$ let $\rho_a \subseteq Dom(a)^{(*)}$. The system $\rho = \{\rho_a | a \in \mathcal{U}\}$ is called the *system of standard relations of data.*

Definition 9.

1. *Let T be a set. A $p-$scheme of T is a system $s = \{S_a \subseteq T^{(*)} | a \in \mathcal{U}\}$ where for all $a \in \mathcal{U}$ there is a mapping $v_a : T \to Dom(a)$ such that:*

$$(t_1, t_2, \ldots, t_n) \in S_a \Longrightarrow (v_a(t_1), v_a(t_2), \ldots, v_a(t_n)) \in \rho_a.$$

2. *For a relation $r = \{t_1, \ldots, t_n\}$ the ρ-scheme of r is $s^r = \{S_a^r | a \in \mathcal{U}\}$ where*

$$(t_{i_1}, t_{i_2}, \ldots, t_{i_m}) \in S_a^r \iff (t_{i_1}[a], t_{i_2}[a], \ldots, t_{i_m}[a]) \in \rho_a.$$

ρ-scheme of a relation is a ρ-scheme.

3. *Let $S_i \subseteq T_i^{(*)}$, $i = 1,2$ and $h : T_1 \to T_2$. We write $S_1 \leq_h S_2$ iff $h(S_1) \subseteq S_2$ where $h(S_1) = \{(h(t_1), \ldots, h(t_m)) | (t_1, \ldots, t_m) \in S_1\}$. Let $s_i = \{S_a^i | a \in U\}$ be ρ-scheme of T_i, $i = 1,2$, respectively, and $h : T_1 \to T_2$. We write $s_1 \leq_h s_2$ (and $h : s_1 \xrightarrow{*} s_2$) iff $S_a^1 \leq_h S_a^2$ for all $a \in \mathcal{U}$. We write $s_1 \leq_* s_2$ if $s_1 \leq_h s_2$ for some $h : T_1 \to T_2$. It is easy to see that if $h : s_1 \xrightarrow{*} s_2$, $k : s_2 \xrightarrow{*} s_3$ then $k \circ h : s_1 \xrightarrow{*} s_3$.*

Similarly to (1) of Proposition 1 it is easy to see that for each ρ-scheme s there is a relation r such that $s^r = s$.

7.3. Structure of the universes of data. We consider the universe of data with standard relations of data $\rho = \{\rho_a | a \in \mathcal{U}\}$ where $\rho_a \subseteq Dom(a)^{(*)}$ for $a \in \mathcal{U}$. In practice we usually have to deal with the following structure:

1. *Trivial structures: cases are possible:*

a. $\rho_a = \emptyset$

b. $\rho_a = Dom(a)^{(*)}$

c. $\rho_a = \{(x) | x \in Dom(a)\}$

2. *Equational structures: Let $=_a$ be an equality relationship in $Dom(a)$. Then $\rho_a = \{(x, y) | x, y \in Dom(a) : x =_a y\}$. In practice we often meet with equational structures of data in the form of an equality, an identity or an equivalence.*

3. *Inequational structure: Let $\leq_a$ be a partial order in $Dom(a)$. Then $\rho_a = \{(x, y) | x, y \in Dom(a) : x \leq_a y\}$. In practice, inequational structure are necessary for those data where we would like to apply, for instance, SORT, INDEX, etc. operations.*

4. *Algebraic structure:* Let p_a be an n-ary operation in $Dom(a)$. Then let ρ_a be the "graph" of p_a, i.e. $\rho_a = \{(x_1, \ldots, x_n, y) | x_1, \ldots, x_n, y \in Dom(a) : p_a(x_1, \ldots, x_n) = y\}$. The algebraic structures can represent logical, numerical or character operations.

8. Conclusion. We have described the advantages of using the schemes in the study of relational databases. The schemes represent more clearly the interactions between the dependencies of data, the dependencies of tuples and the dependencies of attributes. However, this approach is valuable for only the *typed* databases [2, 9] where the relationships of data range within the domain of attributes. Another problem is to find those schemes which are suitable to represent the databases with several structures on each domain.

REFERENCES

[1] C. Beeri, M. Dowd, R. Fagin, R. Statman, On the Structure of Armstrong relations for Functional Dependenies, *J. ACM.* **31** (1984), 30-46.

[2] C. BEERI, M. I. VARDI, Formal systems for tuple and equality generating dependencies, *SIAM J. Comput.* **13** (1984), 76-98.

[3] J. BERMAN, J. BLOCK, Positive Boolean dependencies, *Inform. Process. Letters,* **27** (1988), 147-150.

[4] S. S. COSMADAKIS, P. C. KANELLAKIS, N. SPYRATOS, Partition Semantics for relations, *JCSS,* **33** (1986), 203-233.

[5] J. DEMETROVICS, G. GYEPESI, Some generalized type functional dependencies formalized as equality set of matrices, *Discrete Applied Mathematics* **6** (1983), 35-47.

[6] J. DEMETROVICS, L. RÓNYAI, HUA NAM SON, On the representation of dependencies by propositional logic, Proceedings of MFDBS 91, LNCS 495, Springer-Verlag 1991, 230-242.

[7] J. DEMETROVICS, L. RÓNYAI, HUA NAM SON, Functional dependencies among BDs, (to appear).

[8] J. DEMETROVICS, L. RÓNYAI, HUA NAM SON, On the composition and decomposition of tuples and attributes, Database Theory – Proceedings of ICDT 92, LNCS 646, Springer-Verlag 1992, 71-85.

[9] M. H. GRAHAM, K. WANG, On the equivalence of an Egd to a set of Fd's, *J. ACM,* **37** (1990), 474-490.

[10] G. GOTTLOB, L. LIBKIN, Investigation on Armstrong relations, dependency inference, and excluded functional dependencies, *Acta Cybernetica Hungary,* **9** (1990), 385-402.

[11] H. MANNILA, K.-J. RÄIHÄ, Dependency inference, Proceedings of the 13th VLDB, 1987, 155-158.

[12] J. D. ULLMAN, Principles of database and knowledge-base system, **1**, Computer Science Press, 1988.

Computer and Automation Institute
Hungarian Academy of Sciences
P.O.Box 63, Budapest, H-1502
HUNGARY