Table Algebra of Infinite Tables, Multiset Table Algebra, and their Relationship

Iryna Lysenko

Nizhyn Gogol State University, Grafska Str. 2, 16600 Nizhyn, Ukraine
iryna.glushko@ndu.edu.ua

Abstract. The paper is focused on some theoretical questions of the table databases. Two mathematical formalisms such as table algebra of infinite tables and multiset table algebra are considered. Basic definitions referring to these formalisms are given. This paper also addresses the issue of the relationship between table algebra of infinite tables and multiset table algebra. It is proved that table algebra of infinite tables is not a subalgebra of multiset table algebra since it is not closed in relation to some signature operations of multiset table algebra. These signature operations are determined.

Keywords: Relation databases, table algebra of infinite tables, multiset table algebra.

I. INTRODUCTION

Table algebra of infinite tables generalizes table algebra which was suggested by Redko V., Brona J., Buy D., Poliakov S. in monograph [1]. However, the analysis of the proofs of theorems shows that the finiteness of tables is not used in many cases. That is why in monograph [2] it was done a generalization which means under relation any set of tuples, in particular, infinite. In addition, every table correlated to a certain scheme. A table is a pair, where the first component is an arbitrary set and the second component is a scheme of the table.

Multiset table algebra is also introduced in monograph [2]. In this case, under relation, we understand a multiset, in particular, infinite.

The purpose of this work is to determine is whether table algebra of infinite tables a subalgebra of multiset table algebra.

II. FORMAL DEFINITION OF TABLE ALGEBRA OF INFINITE TABLES

It is necessary here to clarify exactly what is meant by table algebra of infinite tables. Let's introduce the basic concepts of table algebra of infinite tables in terms of monograph [2].

Let A be the set of attributes and D be the universal domain. An arbitrary finite set of attributes $R \subseteq A$ is called a scheme. The tuple of scheme R is the nominal set on pair R, D. The projection of this nominal set for the first component is equal to R.

The table of scheme R ($R \subseteq A$) is a pair $\langle t, R \rangle$, where t is a set (in particular infinite) of tuples of fixed scheme R. The notation $(\langle t, R \rangle)_1$ denotes the first component of the pair $\langle t, R \rangle$, i.e the set t.

We will designate a set of all tuples (tables) of the scheme R by $S(R)$ ($T(R)$ respectively), and the set of all tuples (tables) by S (T respectively). Hence,

$$S = \bigcup_{R \subseteq A} S(R), \quad T(R) = \{ \langle t, R \rangle \mid t \in P(S(R)) \}, \quad T = \bigcup_{R \subseteq A} T(R),$$

where $P(X)$ is a power set of a set X.

Under the table algebra of infinite tables is generally understood to mean an algebra $\langle T, \Omega_{P,\Xi} \rangle$, where T is the set of all tables, $\Omega_{P,\Xi} = \{ \bigcup_R, \bigcap_R, \setminus_R, \sigma_{p,R}, \pi_{X,R}, \underset{R_1,R_2}{\otimes}, \div_{R_2}^{R_1}, Rt_{\xi,R}, \sim_R \}$ is a signature, $p \in P, \xi \in \Xi$, $X, R, R_1, R_2 \subseteq A$, P, Ξ are the sets of parameters. The signature $\Omega_{P,\Xi}$ contains set-theoretic operations (union $\bigcup_R$, intersection $\bigcap_R$, difference $\setminus_R$) and special operations (selection $\sigma_{p,R}$, projection $\pi_{X,R}$, join $\underset{R_1,R_2}{\otimes}$, division $\div_{R_2}^{R_1}$, renaming $Rt_{\xi,R}$, and active complement $\sim_R$).

III. FORMAL DEFINITION OF THE MULTISET TABLE ALGEBRA

Multiset table algebra is multisets analog of the table algebra. In this case, the concept of the table is specified, using concept of the multisets (or bags).

Let's introduce the basic concepts of multiset table algebra in terms of monograph [2].

As before, A is the set of attributes and D is the universal domain. An arbitrary (finite) set of attributes $R \subseteq A$ is called the scheme. A tuple of the scheme R is a nominal set on pair R, D. The projection of this nominal set for the first component is equal to R. The set of all tuples on scheme R is designated as $S(R)$ and the set of all tuples is designated as S.

The table is pair $\langle \psi, R \rangle$, where the first component ψ is an arbitrary multiset, basis of which $\Theta(\psi)$ is the set of tuples of the same scheme and the second component R is a scheme of the table. The set of all table on scheme R is designated as $\Psi(R)$ and the set of all table is designated as $\Psi = \bigcup_R \Psi(R)$.

The notation $Occ(s, \psi)$ denotes the number of duplicate tuple s in the multiset ψ. Let's agree a multiset to write down as $\{s_1^{n_1}, \dots, s_k^{n_k}\}$, where $n_i = Occ(s_i, \psi)$, $i = 1, \dots k$, and $\Theta(\psi) = \{s_1, \dots, s_k\}$ is a of the multiset ψ.

The multiset table algebra is an algebra $\langle \Psi, \Omega_{P,\Xi} \rangle$, where Ψ is the set of all tables, $\Omega_{P,\Xi} = \{\bigcup_{All}^R, \bigcap_{All}^R, \setminus_{All}^R, \sigma_{p,R}, \pi_{X,R}, \underset{R_1,R_2}{\otimes}, Rt_{\xi,R}, \sim_R\}_{X,R,R_1,R_2 \subseteq A}^{p \in P, \xi \in \Xi}$ is the signature, P, Ξ are the sets of parameters.

Let's introduce the basic concepts of multisets in terms of papers [1, 3]. A multiset α with basis U is a function $\alpha : U \to N^+$, where U is an arbitrary set, $N = \{1, 2, \dots\}$ is the set of natural numbers.

Let D be the universe of elements of multiset bases, and then power set $P(D)$ is the universe of multiset bases. Let α be a multiset with basis $U_\alpha = dom\,\alpha$. Here $dom\,\alpha$ is the range of definition of multiset as a function.

A characteristic function of multiset α is a function $\chi_\alpha : D \to N$, the values of which are specified by the following piecewise schema:

$$\chi_\alpha(d) = \begin{cases} \alpha(d) \text{ if } d \in dom\,\alpha, \\ 0, \text{ else;} \end{cases}$$

for all $d \in D$.

The empty multiset $\varnothing_m$ can be defined as a multiset which basis is the empty set.

The 1-multisets are multisets whose range of values is the empty set or a single-element set $\{1\}$. These multisets are the analogues of ordinary sets.

The operations over multisets are defined in terms of characteristic functions in monograph [1]. Authors define operations of multiset union $\bigcup_1$, intersection $\bigcap_1$, difference $\setminus_1$, which build 1-multisets, and operations of multiset union $\bigcup_{All}$, intersection $\bigcap_{All}$, difference $\setminus_{All}$, which build multisets of general view. The Cartesian product of multiset $\otimes$, the operation $Dist(\alpha)$, which build 1-multiset, and analog of a full image for multisets are defined too.

IV. ABOUT RELATIONSHIP BETWEEN TABLE ALGEBRA OF INFINITE TABLES AND MULTISET TABLE ALGEBRA

The 1-multisets are the analogues of ordinary sets. According to this statement table algebra of infinite tables seems to be a subalgebra of multiset table algebra. Let's check it.

For this, suppose that a table of table algebra of infinite tables is a pair $\langle t^1, R \rangle$, where t^1 is a 1-multiset (in particular infinite) of tuples of the fixed scheme R. In this case, obviously, that the set of all tables of table algebra of infinite tables T is a subset of the set of all tables of multiset table algebra Ψ, i.e. $T \subseteq \Psi$.

Now we find out whether the set T is closed with respect to every operation of the signature of multiset table algebra $\Omega_{P,\Xi}$ [4].

Let's start with the set-theoretic operations: union $\bigcup_{All}^R$, intersection $\bigcap_{All}^R$ and difference $\setminus_{All}^R$. Let us consider each operation separately.

Bases of 1-multisets t_1^1 and t_2^1 are designated as $\Theta(t_1^1)$, $\Theta(t_2^1)$ respectively.

The union defined as $\langle t_1^1, R \rangle \bigcup_{All}^R \langle t_2^1, R \rangle = \langle t_1^1 \bigcup_{All} t_2^1, R \rangle$, where $\langle t_1^1, R \rangle$, $\langle t_2^1, R \rangle \in T(R)$.

Basis of the multiset $t_1^1 \bigcup_{All} t_2^1$ of the resulting table is equal to union of bases of multisets of input tables:

$$\Theta(t_1^1 \bigcup_{All} t_2^1) = \Theta(t_1^1) \bigcup \Theta(t_2^1).$$

Duplicate tuples, which appear after implementation of operation, are not removed from the result. The number of duplicates is given by the following formula:

$$Occ(s, t_1^1 \bigcup_{All} t_2^1) = \begin{cases} 1, \text{ if } s \in \Theta(t_1^1) \setminus \Theta(t_2^1) \text{ or } s \in \Theta(t_2^1) \setminus \Theta(t_1^1), \\ 2, \text{ if } s \in \Theta(t_1^1) \bigcap \Theta(t_2^1); \end{cases}$$

where $s \in \Theta(t_1^1) \bigcup \Theta(t_2^1)$. As a result $t_1^1 \bigcup_{All} t_2^1$ is a multiset and it is not a 1-multiset, as we expected.

Example 1. The tables $\langle t_1^1, R \rangle$ and $\langle t_2^1, R \rangle$ are shown in Table 1 and Table 2 respectively, where $R = \{A, B, C\}$. For tuple $s = \{\langle A, a \rangle, \langle B, a \rangle, \langle C, a \rangle\}$ the number of duplicates is 2, $Occ(s, t_1^1 \bigcup_{All} t_2^1) = 2$.

Table 1. The table $\left\langle t_1^1, R \right\rangle$

A	B	C
a	a	a
a	b	b
a	a	b

Table 2. The table $\left\langle t_2^1, R \right\rangle$

A	B	C
a	b	a
a	b	b
b	b	b
b	a	a

The result of union $\left\langle t_1^1, R \right\rangle \bigcup_{All}^R \left\langle t_2^1, R \right\rangle$ we can see in Table 3.

Table 3. The result of the union $t_1^1 \bigcup_{All} t_2^1$

A	B	C
a	a	a
a	b	b
a	a	b
a	b	b
a	b	a
b	b	b
b	a	a

Thus, the set of all table of table algebra of infinite tables $\boldsymbol{T}$ is not closed with respect to the union $\bigcup_{All}^R$.

Find out whether the set $\boldsymbol{T}$ is closed with respect to other operations of the signature of multiset table algebra $\Omega_{P,\Xi}$.

The set $\boldsymbol{T}$ is closed with respect to intersection: $\bigcap_{All}^R : \boldsymbol{T}(R) \times \boldsymbol{T}(R) \to \boldsymbol{T}(R)$, and $\left\langle t_1^1, R \right\rangle \bigcap_{All}^R \left\langle t_2^1, R \right\rangle = \left\langle t_1^1 \bigcap_{All} t_2^1, R \right\rangle$, where $\left\langle t_1^1, R \right\rangle$, $\left\langle t_2^1, R \right\rangle \in \boldsymbol{T}(R)$.

Basis of the multiset $t_1^1 \bigcap_{All} t_2^1$ is equal to intersection of bases of 1-multisets of input tables:
$$\Theta(t_1^1 \bigcap_{All} t_2^1) = \Theta(t_1^1) \bigcap \Theta(t_2^1),$$
and the duplicates are counted as
$$Occ(s, t_1^1 \bigcap_{All} t_2^1) = \min\big(Occ(s, t_1^1), Occ(s, t_2^1)\big) = 1,$$
where $s \in \Theta(t_1^1) \bigcap \Theta(t_2^1)$.

Every tuple, which belongs to $\left\langle t_1^1, R \right\rangle$ and $\left\langle t_2^1, R \right\rangle$ simultaneously has number of duplicates, is equal to 1. The minimum number of this tuple's duplicates in the table $\left\langle t_1^1 \bigcap_{All} t_2^1, R \right\rangle$ is also equal to 1.

Thus, $t_1^1 \bigcap_{All} t_2^1$ is 1-multisets and $\left\langle t_1^1 \bigcap_{All} t_2^1, R \right\rangle \in \boldsymbol{T}(R)$, it means that the result of the intersection of two tables from the set $\boldsymbol{T}$ belongs to the set $\boldsymbol{T}$. Thus, the set $\boldsymbol{T}$ is closed with respect to the intersection $\bigcap_{All}$.

Let us consider the difference operation $\backslash_{All}^R$:
$$\left\langle t_1^1, R \right\rangle \backslash_{All}^R \left\langle t_2^1, R \right\rangle = \left\langle t_1^1 \backslash_{All} t_2^1, R \right\rangle.$$

Basis of the multiset $t_1^1 \backslash_{All} t_2^1$ is defined as $\Theta(t_1^1 \backslash_{All} t_2^1) = \Theta(t_1^1) \backslash \Theta(t_2^1)$.

The number of duplicates is given by the following formula:
$$Occ(s, t_1^1 \backslash_{All} t_2^1) = \begin{cases} 1, & if \ s \in \Theta(t_1^1) \backslash \Theta(t_2^1), \\ 0, & if \ s \in \Theta(t_2^1) \backslash \Theta(t_1^1) \end{cases} = Occ(s, t_1^1) \dot{-} Occ(s, t_2^1)$$
,
where $s \in \Theta(t_1^1) \backslash \Theta(t_2^1)$.

As a result, we get that $t_1^1 \backslash_{All} t_2^1$ is a 1- multisets and $\left\langle t_1^1 \backslash_{All} t_2^1, R \right\rangle \in \boldsymbol{T}(R)$. Consequently, the set $\boldsymbol{T}$ is closed with respect to the difference $\backslash_{All}^R$.

Let $p : S \xrightarrow{\sim} \{true, false\}$ be partial predicate on the set of all tuples. Let us find out whether the set $\boldsymbol{T}$ is closed with respect to the selection over a predicate p. Then $\sigma_{p,R}\big(\left\langle t^1, R \right\rangle\big) = \left\langle t', R \right\rangle$, where $\left\langle t^1, R \right\rangle \in \boldsymbol{T}(R)$. The selection on a table $\left\langle t^1, R \right\rangle$ simply applies the selection predicate to each tuple in $\left\langle t^1, R \right\rangle$ and retains in the result those tuples $s \in \left\langle t^1, R \right\rangle$ for which the predicate evaluates to true.

Basis of multiset of the resulting table is defined as follows:
$$\Theta(t') = \{s \mid s \in \Theta(t^1) \wedge p(s) \simeq true\},$$
where $\simeq$ generalized equality (both parts of equality are simultaneously undefined or both defined and equal [5]).

Consequently, depending on the value of p, either all the duplicates of t^1 are selected or none or them, that is
$$Occ(s, t') = Occ(s, t^1) = 1,$$
where $s \in \Theta(t')$. Therefore, t' is a 1-multiset and $\left\langle t', R \right\rangle \in \boldsymbol{T}(R)$. Thus, the set $\boldsymbol{T}$ is closed with respect to the selection $\sigma_{p,R}$.

Let $X \subseteq A$ be a (finite) set of attributes. Let us find out whether the set $\boldsymbol{T}$ is closed with respect to the projection over a set of attributes X. Then $\pi_{X,R}\big(\left\langle t^1, R \right\rangle\big) = \left\langle t', R \bigcap X \right\rangle$, where $\left\langle t^1, R \right\rangle \in \boldsymbol{T}(R)$.

Basis of multiset t' is defined as follows:
$$\Theta(t') = \{s \mid X \mid s \in \Theta(t^1)\}.$$

Duplicate tuples, which appear after implementation of operation, are not removed from the result. Specifically, the number of duplicates is given by the following formula:
$$Occ(s', t') = \sum_{\substack{s \in \Theta(t^1), \\ s \mid X = s'}} Occ(s, t^1)$$
,

where $s' \in \Theta(t')$. Consequently, t' is not a 1-multiset. Thus, the set of all table of table algebra of infinite tables $\boldsymbol{T}$ is not closed with respect to the projection $\pi_{X,R}$.

Example 2. Let's consider the table $\langle t_1^1, R \rangle$ from Table 1 and find $\pi_{\{A,B\},R}(\langle t^1, R \rangle)$. The result is

$$\pi_{\{A,B\},R}(\langle t^1, R \rangle) = \langle \{\{\langle A,a \rangle, \langle B,b \rangle\}^2, \{\langle A,b \rangle, \langle B,b \rangle\}^1, \{\langle A,a \rangle, \langle B,b \rangle\}^1\}, R_1 \rangle$$

, where $R_1 = \{A, B\}$.

Let us check the join operation $\underset{R_1,R_2}{\otimes}$. Then

$$\langle t_1^1, R_1 \rangle \underset{R_1,R_2}{\otimes} \langle t_2^1, R_2 \rangle = \langle t', R_1 \cup R_2 \rangle, \text{ where } \langle t_1^1, R_1 \rangle \in \boldsymbol{T}(R_1),$$

$\langle t_2^1, R_2 \rangle \in \boldsymbol{T}(R_2)$. In other words, each tuple of $\langle t_1^1, R_1 \rangle$ is paired with each tuple of $\langle t_2^1, R_2 \rangle$, regardless of whether it is a duplicate or not.

Basis of multiset t' is defined as follows:

$$\Theta(\psi') = \{s' | \exists s_1 \exists s_2 (s_1 \in \Theta(\psi_1) \wedge s_2 \in \Theta(\psi_2) \wedge s_1 \approx s_2 \wedge s' = s_1 \cup s_2)\}$$

.

The number of duplicates is given by the following formula:

$$Occ(s', t') = Occ(s' | R_1, t_1^1) \cdot Occ(s' | R_2, t_2^1) = 1,$$

where $s' \in \Theta(t')$.

Thus, t' is a 1-multiset. Consequently, the set $\boldsymbol{T}$ is closed with respect to the join $\underset{R_1,R_2}{\otimes}$.

Let us consider the renaming operation. Then $R\psi_{\xi,R}(\langle t^1, R \rangle) = \langle Rs_\eta[t^1], \eta[R] \rangle$, $t^1 \in T_\xi(R)$, where $\eta = \xi \cup \mathrm{id}_{A \setminus dom\xi}$, $Rs_\eta(s) = \{\langle \eta(A), s(A) \rangle | A \in R\}$, $Rs_\eta[t^1]$ full image of 1-multiset t^1 with basis $\Theta(t^1)$ regarding of function Rs_η.

Basis of multiset $Rs_\eta[t^1]$ is full image of set $\Theta(t^1)$ regarding of function $Rs_{\eta,R}$. The number of duplicates is given by the following formula:

$$Occ(s', Rs_\eta[t^1]) = Occ(s, t^1) = 1,$$

where $s \in Rs_\eta^{-1}(s')$, $s' \in \Theta(Rs_\eta[t^1])$.

In particular, the number of occurrences of each tuple in the result does not change, because the function Rs_η is injective. Thus, $Rs_\eta[t^1]$ is a 1-multiset. Consequently, the set $\boldsymbol{T}$ is closed with respect to the renaming operation.

Finally, let's consider the active complement operation $\sim_R$. Then $\sim_R (\langle t^1, R \rangle) = C(\langle t^1, R \rangle) \setminus_{All}^R \langle t^1, R \rangle$, where the saturation is defined as

$$C(\langle t^1, R \rangle) = \pi_{\{A_1\},R}(\langle t^1, R \rangle) \underset{\{A_1\},\{A_2\}}{\otimes} \cdots \underset{\{A_1,\ldots A_{n-1}\},\{A_n\}}{\otimes} \pi_{\{A_n\},R}(\langle t^1, R \rangle)$$

. It is shown above that the set $\boldsymbol{T}$ is not closed with respect to the projection $\pi_{X,R}$, that is why the set $\boldsymbol{T}$ is also not closed with respect to the active complement $\sim_R$.

Hence, the set of all tables of table algebra of infinite tables is closed with respect to the intersection, difference, selection, join and renaming and the set of all tables of table algebra of infinite tables is not closed with respect to the union, projection and active complement. Consequently, table algebra of infinite tables is not a subalgebra of multiset table algebra.

V. CONCLUSION

In this paper applying the theorem-plural and logical-algebraic methods found that table algebra of infinite tables does not form subalgebra of multiset table algebra since it is not closed with respect to the union, projection and active complement. Thus, multiset table algebra is not a wider formalism then table algebra of infinite tables.

REFERENCES

[1]. Redko, V., Buy, D., Brona J., Polyakov S.: Relational Databases: Table Algebras and SQL-like Language. Publishing house Academperiodica, Kyiv (2001).

[2]. Buy, D., Glushko, I.: Calculi and extensions of table algebras signature. NDU im. M. Gogol, Nizhyn (2016).

[3]. Buy, D., Bogatyreva, J.: Multiset's theory: bibliography, application in table data bases theory. Radioelectronic and computer systems. Journal 7, 56–61 (2010).

[4]. Glushko, I.: About relationship between table algebra of infinite tables and multiset table algebra. In 11th International Conference of Programming on Proceedings, pp.159–163. CEUR Workshop Proceedings (2018), http://ceur-ws.org/Vol-2139/159-163.pdf.

[5]. Cutland, N.: Computability. An introduction to recursive function theory. Myr, Moscow (1983).