A Formal Mathematical Semantics of Advanced Operations of Multiset Table Algebra

Iryna Glushko
Applied Mathematics, Informatics and Educational Measurement Department
Nizhyn Gogol State University
Nizhyn, Ukrain
iryna.glushko@ndu.edu.ua

Abstract—Multiset table algebra is considered. The notion of a table specified using the notion of a multiset (or bag). A signature of multiset table algebra is filled up with new operations such as inner and outer joins, semi-join and aggregate operations. A formal mathematical semantics of these operations is defined. The special element NULL is inserted in the universal domain for a define of the outer join.

Keywords—relation databases; multiset table algebra; inner joins; outer joins; semi-join, aggregate operations

I. INTRODUCTION
There are many applications the most peculiar feature of which is multiplicity and repeatability data. For example, these are sociological polls of different population groups, calculations on DNA and others. Commercial relational database systems are almost invariably based on multisets instead of sets. In other words, tables are in general allowed to include duplicate tuples. For example, the data model of SQL is relational in nature, as well as the relevant operations. However, unlike relational algebra, the tables manipulated by SQL are not relations, but, rather, multisets. The reason for this peculiarity is twofold. First, this is due to a practical reason: since SQL tables may be very large, duplicate elimination might become a bottleneck for the computation of the query result. Second, SQL extends the set of query operators by means of aggregate functions, whose operands are in general required to be multisets of values.

The relational model is based on the sets of tuples, i.e. it does not allow duplicate tuples in a relation [1]. So, naturally there is a need to expand possibilities of relational databases due to use of multisets. This problem was also considered in [2-5]. However this question requires specification and extension because in the specified works the due attention isn’t paid to operations of inner and outer joins, semijoin and aggregate operations of multiset table algebra.

II. MULTISET: BASIC DEFINITIONS
Let’s introduce the basic concepts of multisets in terms of monograph [5].

Definition 1. A multiset \( \alpha \) with basis \( U \) is a function \( \alpha: U \to \mathbb{N}^+ \), where \( U \) is an arbitrary set, \( \mathbb{N}^+ = \{1, 2, \ldots\} \) is the set of natural numbers without zero.

Let \( D \) be a universe of element of multiset bases, and then power set \( P(D) \) – a universe of multiset bases. Let \( \alpha \) be a multiset with basis \( U_\alpha = \text{dom}\alpha \). Here \( \text{dom}\alpha \) is the range of definition of multiset as a function.

Definition 2. A characteristic function of multiset \( \alpha \) is a function \( \chi_\alpha: D \to \mathbb{N} \), the values of which are specified by the following piecewise schema: \( \chi_\alpha(d) = \begin{cases} \alpha(d) & \text{if } d \in \text{dom}\alpha, \\ 0 & \text{else}; \end{cases} \) for all \( d \in D \).

Definition 3. An empty multiset \( \emptyset \) is a multiset a characteristic function of which is a constant function, value of which is everywhere equal zero.

Definition 4. A rank of finite multiset \( \alpha \) is a sum of duplicate elements of its basis \( \|\alpha\| = \sum \alpha(d) \); wherein \( \|\emptyset\| = 0 \).

Let’s introduce a binary relation inclusion over multisets.

Definition 5. Multiset \( \beta \) is included in multiset \( \alpha \) (\( \beta \leq \alpha \)), if \( U_\beta \subseteq U_\alpha \) \& \( \forall d \in U_\beta \Rightarrow \beta(d) \leq \alpha(d) \). Directly from definition follows that this relation is a partial order.

The 1-multisets are the multisets whose range of values is an 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 [5]. There are operations of multiset union $\bigcup_{A}$, intersection $\bigcap_{A}$, difference $\setminus_{A}$, which build multisets of general view. The Cartesian product of multiset $\otimes$, the operation $\text{Dist}(x)$, which build 1-multiset, and analog of a full image for multisets are defined too.

III. MULTISET TABLE ALGEBRA: BASIC DEFINITIONS

Among the two sets that are considered, $A$ is the set of attributes and $D$ is the universal domain.

Definition 6. An arbitrary (finite) set of attributes $R \subseteq A$ is called the scheme.

Definition 7. 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 $S$.

Definition 8. A table is a pair $\langle \psi, R \rangle$, where the first component $\psi$ is an arbitrary multiset basis of which $\Theta(\psi)$ is an arbitrary set (in particular infinite) of tuples of the scheme $R$ and other component $R$ is a scheme of table.

Thus, a certain a scheme is ascribed to every 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 $\text{Occ}(s, \psi)$ denotes the number of duplicate tuple $s$ in the multiset $\psi$. Let’s agree a multiset to write down as $\{s_{1}, ..., s_{n}\}$, where $n = \text{Occ}(s_{i}, \psi)$, $i = 1, ..., k$, and $\Theta(\psi) = \{s_{1}, ..., s_{k}\}$ is a basis of the multiset $\psi$.

Definition 9. The multiset table algebra is the algebra $\langle \Psi, \Omega, \Xi \rangle$, where $\Psi$ is the set of all tables, $\Omega_{p, \Xi} = \{\bigcup_{A_{p}}, \bigcap_{A_{p}}, \setminus_{A_{p}}, \sigma_{p, p}, \pi_{x, r}, \otimes_{R_{1, \Xi}}, \sum_{R_{1, \Xi}} \mid p, \Xi \subseteq A \}$ is the signature, $P$, $\Xi$ are the sets of parameters.

The operations of signature $\Omega_{p, \Xi}$ are defined in [6].

IV. THE ADVANCED OPERATIONS

The advanced operations include inner and outer joins, semijoin, aggregate operations.

A. Inner Joints

There are four kinds of inner join operations Cartesian join, natural join, join using attributes $A_{1}, ..., A_{k}$ and join on predicate $p$. Let’s define them.

Definition 10. The Cartesian Join of table on scheme $R_{1}$ and table on scheme $R_{2}$, moreover $R_{1} \cap R_{2} = \emptyset$, is a binary parametric operation $\text{CJ}$ of the form

$$\text{CJ} : \Psi(R_{1}) \times \Psi(R_{2}) \rightarrow \Psi(R_{1} \cup R_{2})$$

$$\langle \psi_{1}, R_{1} \rangle \text{CJ} \langle \psi_{2}, R_{2} \rangle = \langle \psi', R_{1} \cup R_{2} \rangle$$

where $\langle \psi_{1}, R_{1} \rangle \in \Psi(R_{1})$, $\langle \psi_{2}, R_{2} \rangle \in \Psi(R_{2})$.

The basis of the multiset $\psi'$ is defined by follow:

$$\Theta(\psi') = \{s \mid \exists s_{1} \exists s_{2} (s_{1} \in \Theta(\psi_{1}) \land s_{2} \in \Theta(\psi_{2}) \land s_{1} \setminus s_{2} \subseteq s \} \}$$

The number of duplicates is given by the following formula:

$$\text{Occ}(s, \psi') = \text{Occ}(s, \psi_{1}) \cdot \text{Occ}(s, \psi_{2})$$

where $s \in \Theta(\psi')$ and $s = s_{1} \cup s_{2}$.

Definition 11. The Inner Natural Join of table on scheme $R_{1}$ and table on scheme $R_{2}$ is a binary parametric operation written as $\otimes_{R_{1}, R_{2}}$, whose value is the table on scheme $R_{1} \cup R_{2}$ consisting of all the unions of compatible tuples of input tables. Hence,

$$\otimes_{R_{1}, R_{2}} : \Psi(R_{1}) \times \Psi(R_{2}) \rightarrow \Psi(R_{1} \cup R_{2})$$

$$\langle \psi_{1}, R_{1} \rangle \otimes_{R_{1}, R_{2}} \langle \psi_{2}, R_{2} \rangle = \langle \psi', R_{1} \cup R_{2} \rangle$$

where $\psi_{1} \in \Psi(R_{1})$, $\psi_{2} \in \Psi(R_{2})$.

In other words, each tuple of $\psi_{1}$ is paired with each tuple of $\psi_{2}$, regardless of whether it is a duplicate or not. The basis of the multiset $\psi'$ is defined by follow:

$$\Theta(\psi') = \{s \mid \exists s_{1} \exists s_{2} (s_{1} \in \Theta(\psi_{1}) \land s_{2} \in \Theta(\psi_{2}) \land s_{1} \setminus s_{2} \subseteq s) \}$$

The number of duplicates is given by the following formula:

$$\text{Occ}(s, \psi') = \text{Occ}(s, \psi_{1}) \cdot \text{Occ}(s, \psi_{2})$$

where $s \in \Theta(\psi')$ and $s = s_{1} \cup s_{2}$. The relation $\approx$ is a binary relation of compatibility of tuples $s_{1} \approx s_{2} \Leftrightarrow s_{1} \mid R = s_{2} \mid R$ and $s_{1} \mid R$ is the restrictions of tuple $s_{1}$ on the scheme $R$ [5].

The Inner Join using $A_{1}, ..., A_{k}$ on table on scheme $R_{1}$ and table on scheme $R_{2}$, moreover $R_{1} \cap R_{2} = \emptyset$, is a binary parametric operation of the form

$$\otimes_{A_{1}, ..., A_{k}, R_{1}, R_{2}} : \Psi(R_{1}) \times \Psi(R_{2}) \rightarrow \Psi(R_{1} \cup R_{2})$$

$$\langle \psi_{1}, R_{1} \rangle \otimes_{A_{1}, ..., A_{k}, R_{1}, R_{2}} \langle \psi_{2}, R_{2} \rangle = \langle \psi', R_{1} \cup R_{2} \rangle$$

where $\langle \psi_{1}, R_{1} \rangle \in \Psi(R_{1})$, $\langle \psi_{2}, R_{2} \rangle \in \Psi(R_{2})$.

Moreover all $A_{1}, ..., A_{k}$ are pairwise different, $n \geq 1$, and $R_{1} \cap R_{2} = \emptyset$. If input tables have also other general attributes which differ from $A_{1}, ..., A_{k}$, before join they needs to be renamed.

The basis of the multiset $\psi'$ is defined by follow:

$$\Theta(\psi') = \{s \mid \exists s_{1} \exists s_{2} (s_{1} \in \Theta(\psi_{1}) \land s_{2} \in \Theta(\psi_{2}) \land$$
The number of duplicates is given by the following formula: 
\[ \text{Occ}(s, \psi') = \text{Occ}(s, \psi_s) \], where \( s \in \Theta(\psi') \).

**B. Outer Join**

We can lose information when using the inner join operations because the tuples which are not compatible will not be represented in the output table. The outer join operations use when it is necessary to consider the tuples of input tables which didn’t get to result of the inner join operations.

Let \( \text{NULL} \) be a special element of the universal domain \( D \cdot \text{NULL} \) used to denote absent values in the output table. Let \( s_{R,\text{NULL}} \) be a constant tuple on scheme \( R \), i.e.

\[ \left( \begin{array}{c}
1
\end{array} \right) \]

There is one logical scheme for definition of the outer join operations [5].

Let \( \varphi : \Psi(R_1) \times \Psi(R_2) \rightarrow \Psi(R_1 \cup R_2) \) be some partial binary operation on the set of all tables and 
\[ \left( \begin{array}{c}
\varphi(s_1, \psi_1) \rightarrow \left( \begin{array}{c}
\varphi(s_2, \psi_2)
\end{array} \right)
\end{array} \right) \] for all tables 

\[ \left( \begin{array}{c}
\varphi(s_1, \psi_1) = \left( \begin{array}{c}
\varphi(s_2, \psi_2)
\end{array} \right)
\end{array} \right) \] in \( \text{dom} \varphi \).

Let’s notice that the operations 

\[ \left( \begin{array}{c}
\varphi
\end{array} \right)_{\bigcup R_1 \cup R_2} \quad \left( \begin{array}{c}
\varphi
\end{array} \right)_{\bigcap R_1 \bigcap R_2} \quad \left( \begin{array}{c}
\varphi
\end{array} \right)_{\bigcap R_1 \bigcap R_2} \quad \left( \begin{array}{c}
\varphi
\end{array} \right)_{\bigcup R_1 \bigcup R_2} \]

are such.

We fix two tables \( \left( \begin{array}{c}
\varphi_1, R_1
\end{array} \right) \), \( \left( \begin{array}{c}
\varphi_2, R_2
\end{array} \right) \) from range of definition of the operation \( \varphi \).

Then the table \( \left( \begin{array}{c}
\varphi_1, R_1
\end{array} \right) \) takes the following form

\[ \left( \begin{array}{c}
\varphi_1 \cap \varphi_2, R_1 \bigcup R_2
\end{array} \right) \]
In other words, the tuples of the table \( \psi_1 \cap \psi_2, R \) are used in formation of result of the join operation, and tuples of the table \( \psi_1 \neq \psi_2, R \) are not used.

We obtain a representation of the table \( \psi_2, R \) replacing the roles of the tables \( \psi_1, R \) and \( \psi_2, R \) in the presentation of the table \( \psi_1, R \).

Let’s notice that if the operation \( \varphi \) coincides with the operation \( \otimes \) then the table \( \psi_1 \cap \psi_2, R \) is the semi-join of the tables \( \psi_1, R \) and \( \psi_2, R \), i.e.

\[
\psi_1 \cap \psi_2, R = \psi_1, R \otimes \psi_2, R.
\]

There are four kinds of the outer joins operations which are induced of the inner join operation \( \varphi \): outer left join, outer right join, outer full join and union join. Let’s define them.

Consider the following inner natural joins

\[
\psi_1 \cap \psi_2, R \otimes \{\psi_2_{\mathrm{R.,null}} \cup \{R \backslash R \} \} = \psi_1, R \cup \psi_2, R,
\]

where

\[
\Theta(\psi^{'}) = \{s_1 \cup s_2_{\mathrm{R.,null}} \mid s_1 \in \Theta(\psi_1 \backslash \psi_2)\},
\]

\[
\text{Occ}(s', \psi^{'}) = \text{Occ}(s_1, \psi_1 \backslash \psi_2), \quad s' \in \Theta(\psi^{'})
\]

and

\[
\psi_1 \cap \psi_2, R \otimes \{\psi_2_{\mathrm{R.,null}} \cup \{R \backslash R \} \} = \psi_1, R \cup \psi_2, R,
\]

where

\[
\Theta(\psi^{'}) = \{s_1 \cup s_2_{\mathrm{R.,null}} \mid s_1 \in \Theta(\psi_1 \backslash \psi_2)\},
\]

\[
\text{Occ}(s', \psi^{'}) = \text{Occ}(s_1, \psi_1 \backslash \psi_2), \quad s' \in \Theta(\psi^{'})
\]

Definition 14. The Outer Left Join operation is a partial binary operation of the form \( \varphi_1 : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2) \), where dom \( \varphi_1 = \text{dom } \varphi \) and

\[
\varphi_1(\psi_1, R_1), \psi_2, R_2) = \varphi(\psi_1, R_1), (\psi_2, R_2) \bigcup^{\psi_1, R_1} R_2.
\]

Definition 15. The Outer Right Join operation is a partial binary operation of the form \( \varphi_2 : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2) \), where dom \( \varphi_2 = \text{dom } \varphi \) and

\[
\varphi_2(\psi_1, R_1), \psi_2, R_2) = \varphi(\psi_1, R_1), (\psi_2, R_2) \bigcup^{\psi_1, R_1} R_2.
\]

Definition 16. The Outer Full Join operation is a partial binary operation of the form \( \varphi_f : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2) \), where dom \( \varphi_f = \text{dom } \varphi \) and

\[
\varphi_f(\psi_1, R_1), \psi_2, R_2) = \varphi(\psi_1, R_1), (\psi_2, R_2) \bigcup^{\psi_1, R_1} R_2.
\]

Definition 17. The Outer Union Join operation is a partial binary operation of the form \( \varphi_j : \Psi(R_1) \times \Psi(R_2) \to \Psi(R_1 \cup R_2) \), where dom \( \varphi_j = \text{dom } \varphi \) and

\[
\varphi_j(\psi_1, R_1), (\psi_2, R_2) = \varphi(\psi_1, R_1), (\psi_2, R_2) \bigcup^{\psi_1, R_1} R_2.
\]

C. Aggregate Operations

The five types of aggregate operations discussed in this article are SUM, AVERAGE, MAXIMUM, MINIMUM, COUNT. The aggregate operations transform a finite table into a table with single tuple and single attribute.

Consider the table \( \psi, R \in \Psi(R) \), where \( \psi \) is a finite multiset and \( A \in R \). Let \( \alpha_\psi \) be a multiset of column with attribute \( A \) of table \( \psi, R \) which contains all elements including duplicates.

Then

\[
\alpha(\psi) = \{d \mid \exists s \in \Theta(\psi) \land \{A, d\} \in s\}
\]

is an analogue of active domain of the attribute \( A \) [5]. The number of duplicates of element \( d \in \Theta(\psi) \) is given by the following formula:

\[
\alpha(d) = \text{Occ}([\{A, d\} \cdot \{\pi_{1(A)}, \{\psi, R\}\}]) = \sum_{s \in \Theta^{|s|} \psi, A} \text{Occ}(s, \psi).
\]

Let \( 2^D \) be a family of all multisets, bases of which are the finite subsets of the set \( D \). Here \( D' \subseteq D \) is a subset of the universal domain.

Let \( \text{Num} \) is a numerical subset of the universal domain \( D \) that is closed under addition. Extend the set \( \text{Num} \) by the special element \( \text{NULL} \). We will not extend the operation of addition to the case where at least one of the arguments is \( \text{NULL} \).

Let’s define the aggregate operations. At first the five aggregate functions – count, sum, average, maximum, minimum – are defined on a finite multiset and then these functions are transferred to the tables.

Definition 18. The aggregate operation \( \text{Sum}_{\alpha_\psi} \) by the attribute \( A \) of the finite table on scheme \( R \), \( A \in R \), is a unary parametric operation of the form

\[
\text{Sum}_{\alpha_\psi} : \Psi(R) \to \Psi([A]).
\]
$\langle \psi, R \rangle \in \Psi(R)$. The $\text{Sum}(\alpha_A)$ function is applied to a column with attribute $A$ in the table $\langle \psi, R \rangle$, the result obtained is the sum of every value occurrence in $\alpha_A$. In addition, NULL values don’t undertake in attention and it is assumed that the column contains only data of numeric type.

Thus, $\text{Sum} : 2^\text{Num} \rightarrow \text{Num}$,

$$\text{Sum}(\alpha_A) = \begin{cases} \text{NULL} & \text{if } \Theta(\alpha_A) = \emptyset; \\ \text{NULL} & \text{if } \Theta(\alpha_A) = \{ \text{NULL} \}; \\ \sum_{d \in \Theta(\alpha_A)} (d) & \text{if } \Theta(\alpha_A) \neq \emptyset. \end{cases}$$

So, we have $\text{Sum}(\{ \text{NULL} \}) = \text{NULL}$, $\text{Sum}(\{d_1^n, \ldots, d_k^n\}) = \sum_{i=1}^k d_in_i$ if all elements $d_i$, $i = 1, k$, differ from NULL.

In the case of the empty table $\langle \emptyset, R \rangle$ we have $\text{Sum}_{A,R}(\langle \emptyset, R \rangle) = \langle \{ (A, \emptyset) \} \rangle \{ \{ A \} \}$, here $\emptyset_R = \emptyset_m$.

Example 1. Let $\langle \psi, R \rangle$ be the table of Fig 1. Then $\text{Sum}_{A,R}(\langle \psi, R \rangle) = \langle \{ (A, \emptyset) \} \rangle \{ \{ A \} \}$, $\text{Sum}_{B,R}(\langle \psi, R \rangle) = \langle \{ (B, \emptyset) \} \rangle \{ \{ B \} \}$, $\text{Sum}_{C,R}(\langle \psi, R \rangle) = \langle \{ (C, \emptyset) \} \rangle \{ \{ C \} \}$. In Fig. 2, Fig. 3 and Fig. 4 reactivly, we see the tables $\text{Sum}_{A,R}(\langle \psi, R \rangle)$, $\text{Sum}_{B,R}(\langle \psi, R \rangle)$, $\text{Sum}_{C,R}(\langle \psi, R \rangle)$.

![Fig. 1. Table $\langle \psi, R \rangle$](image1)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>NULL</td>
</tr>
</tbody>
</table>

![Fig. 2. Table $\text{Sum}_{A,R}(\langle \psi, R \rangle)$](image2)

$\text{Sum}_{A,R}(\langle \psi, R \rangle)$

![Fig. 3. Table $\text{Sum}_{B,R}(\langle \psi, R \rangle)$](image3)

$\text{Sum}_{B,R}(\langle \psi, R \rangle)$

![Fig. 4. Table $\text{Sum}_{C,R}(\langle \psi, R \rangle)$](image4)

$\text{Sum}_{C,R}(\langle \psi, R \rangle)$

Let $\leq$ be a linear order on the universal domain $D$.

Definition 19. The aggregate operation $\text{Min}_{A,R}$ by the attribute $A$ of the finite table on scheme $R$, $A \in R$, is a unary parametric operation of the form $\text{Min}_{A,R} : \Psi(R) \rightarrow \Psi(\{A\})$,

$$\text{Min}_{A,R}(\langle \psi, R \rangle) = \langle \{ (A, \text{Min}(\alpha_A)) \} \rangle \{ \{ A \} \},$$

where $\langle \psi, R \rangle \in \Psi(R)$. The $\text{Min}(\alpha_A)$ function is applied to a column with attribute $A$ in the table $\langle \psi, R \rangle$, the result obtained is the minimum value among values of $\alpha_A$. In addition, NULL values don’t undertake in attention.

Thus, $\text{Min} : 2^\text{Num} \rightarrow D$,

$$\text{Min}(\alpha_A) = \begin{cases} \text{NULL} & \text{if } \Theta(\alpha_A) = \emptyset; \\ \text{NULL} & \text{if } \Theta(\alpha_A) = \{ \text{NULL} \}; \\ \min_{d \in \Theta(\alpha_A)} (d) & \text{if } \Theta(\alpha_A) \neq \emptyset. \end{cases}$$

We have $\text{Min}(\emptyset) = \text{NULL}$. $\text{Min}(\{ \text{NULL} \}) = \text{NULL}$, $\text{Min}(\{d_1^n, \ldots, d_k^n\}) = \min(d_1, \ldots, d_k)$ if all elements $d_i$, $i = 1, k$, differ from NULL.

In the case of the empty table $\langle \emptyset, R \rangle$ we have $\\text{Min}_{A,R}(\langle \emptyset, R \rangle) = \langle \{ (A, \emptyset) \} \rangle \{ \{ A \} \}$, here $\emptyset_R = \emptyset_m$.

Example 2. Let $\langle \psi, R \rangle$ be the table of Fig 1. Then $\text{Min}_{A,R}(\langle \psi, R \rangle) = \langle \{ (A, \emptyset) \} \rangle \{ \{ A \} \}$, $\text{Min}_{B,R}(\langle \psi, R \rangle) = \langle \{ (B, \emptyset) \} \rangle \{ \{ B \} \}$, $\text{Min}_{C,R}(\langle \psi, R \rangle) = \langle \{ (C, \emptyset) \} \rangle \{ \{ C \} \}$.

Definition 20. The aggregate operation $\text{Max}_{A,R}$ by the attribute $A$ of the finite table on scheme $R$, $A \in R$, is a unary parametric operation of the form $\text{Max}_{A,R} : \Psi(R) \rightarrow \Psi(\{A\})$,

$$\text{Max}_{A,R}(\langle \psi, R \rangle) = \langle \{ (A, \text{Max}(\alpha_A)) \} \rangle \{ \{ A \} \},$$

where $\langle \psi, R \rangle \in \Psi(R)$. The $\text{Max}(\alpha_A)$ function is applied to a column with attribute $A$ in the table $\langle \psi, R \rangle$, the result obtained is the maximum value among values of $\alpha_A$. In addition, NULL values don’t undertake in attention.
Thus,\( \text{Max} : 2^D \rightarrow D \).

\[
\text{Max}(\alpha) = \begin{cases} 
\text{NULL} \text{ if } \Theta(\alpha) = \emptyset; \\
\text{NULL} \text{ if } \Theta(\alpha) = \{\text{NULL}\}; \\
\max\{d \in \Theta(\alpha) \setminus \{\text{NULL}\} \mid \{\text{NULL}\} \notin \Omega \} \text{ if } \Theta(\alpha) \setminus \{\text{NULL}\} \neq \emptyset.
\end{cases}
\]

We have \( \text{Max}(\emptyset) = \text{NULL}, \text{Max}(\{\text{NULL}\}) = \text{NULL}, \text{Max}(\{d_1, \ldots, d_n\}) = \max\{d_1, \ldots, d_n\} \) if all elements \( d_i, i = 1, k \), differ from \( \text{NULL} \).

In the case of the empty table \( \langle y \subseteq R \rangle \) we have
\[
\text{Max}_{A,R}(\langle y \subseteq R \rangle) = \{\langle A, \text{NULL} \rangle \} \setminus \{A\},
\]

where \( y \subseteq R \). The \( \text{Max}(\alpha) \) function is applied to a column with attribute \( A \) in the table \( \langle y \subseteq R \rangle \), the result obtained is the count of all values of \( A \) which differ from \( \text{NULL} \).

Example 3. Let \( \langle y, R \rangle \) be the table of Fig 1. Then
\[
\text{Max}_{A,R}(\langle y, R \rangle) = \{\langle A,2 \rangle \} \setminus \{A\},
\]
\[
\text{Max}_{B,R}(\langle y, R \rangle) = \{\langle B,3 \rangle \} \setminus \{B\},
\]
\[
\text{Max}_{C,R}(\langle y, R \rangle) = \{\langle C,3 \rangle \} \setminus \{C\}.
\]

Definition 21. The aggregate operation \( \text{Count}_{A,R} \) by the attribute \( A \) of the finite table on scheme \( R, A \in R \), is a unary parametric operation of the form
\[
\text{Count}_{A,R} : \Psi(\langle y \rangle) 
\]
\[
\text{Count}_{A,R}(\langle y, R \rangle) = \{\langle A,\text{Count}_{A}(\alpha) \rangle \} \setminus \{A\},
\]

where \( y \in \Psi(\langle y \rangle) \). The \( \text{Count}(\alpha) \) function is applied to a column with attribute \( A \) in the table \( \langle y, R \rangle \), the result obtained is the arithmetic mean of values in \( A \) which differ from \( \text{NULL} \).

Thus, \( \text{Avg} : 2^D \rightarrow N \) and \( \text{Count}(\alpha) = \frac{\text{Sum}(\alpha)}{\text{Count}(\alpha)} \).

We have
\[
\text{avg}(\emptyset) = \text{null} \quad \text{and} \quad \text{Count}(\alpha) = \null,
\]
\[
\text{avg}(\{\text{NULL}\}) = \null \quad \text{and} \quad \text{Count}(\alpha) = \null,
\]
\[
\text{avg}(\{d_1, \ldots, d_n\}) = \frac{\text{Sum}(\{d_1, \ldots, d_n\})}{\text{Count}(\{d_1, \ldots, d_n\})} = \frac{\sum_{i=1}^{n} d_i}{n_1 + \ldots + n_k} \quad \text{if all elements } d_i, i = 1, k, \text{ differ from NULL.}
\]

In the case of the empty table \( \langle y \subseteq R \rangle \) we have
\[
\text{Avg}_{A,R}(\langle y \subseteq R \rangle) = \{\langle A, \text{NULL} \rangle \} \setminus \{A\},
\]

Example 5. Let \( \langle y, R \rangle \) be the table of Fig 1. Then
\[
\text{Avg}_{A,R}(\langle y, R \rangle) = \{\langle A,2 \rangle \} \setminus \{A\},
\]
\[
\text{Avg}_{B,R}(\langle y, R \rangle) = \{\langle B,6 \rangle \} \setminus \{B\},
\]
\[
\text{Avg}_{C,R}(\langle y, R \rangle) = \{\langle C,6 \rangle \} \setminus \{C\}.
\]

Definition 23. The aggregate operation \( \text{Count}_{A,R}(\bullet) \) by the attribute \( A \) of the finite table on scheme \( R, A \in R \), is a unary parametric operation of the form
\[
\text{Count}_{A,R}(\bullet) : \Psi(\langle y \rangle) 
\]
\[
\text{Count}_{A,R}(\bullet)(\langle y, R \rangle) = \{\langle A,\text{Count}_{A}(\alpha) \rangle \} \setminus \{A\},
\]

where \( y \in \Psi(\langle y \rangle) \), and \( \| y \| \) is the rank of the multiset \( y \).
The operation \( \text{Count}_{A,R}(\star) \) finds the number of tuples in the table \( \langle \psi, R \rangle \).

In the case of an empty table \( \langle \psi, R \rangle \) we have
\[
\text{Count}_{A,R}(\star)(\langle \psi, R \rangle) = \{\{\langle A,0 \rangle\}, \{A\}\},
\]
\[
\psi = \emptyset.
\]

Example 5. Let \( \langle \psi, R \rangle \) be the table of Fig 1. Then
\[
\text{Count}_{A,R}(\star)(\langle \psi, R \rangle) = \{\{\langle A,3 \rangle\}, \{A\}\},
\]
\[
\text{Count}_{B,R}(\star)(\langle \psi, R \rangle) = \{\{\langle B,5 \rangle\}, \{B\}\},
\]
\[
\text{Count}_{C,R}(\star)(\langle \psi, R \rangle) = \{\{\langle C,3 \rangle\}, \{C\}\}.
\]

V. CONCLUSIONS

In this paper the multiset table algebra is considered. The signature of the multiset table algebra is filled up with new operations such as inner and outer join, semijoin and aggregate operations. The special element NULL is inserted in the universal domain for a define of outer operations.

It should also be noted that a parameter of aggregate operations is not necessarily only a single attribute; it also can be some function of the tuples.


