IVC

July 28, 2011

1 Definitions

In this draft we want to explain the notion of domain. The domain of a variable is the set of values that a variable can take. The domain of all the variables in a query expression can easily be created recursively if some rules will be respected. We will use through out the entire paper the notation of $\mathsf{dom}_{\vec{x}}(q)$ for the domain of a set of variables, where $\vec{x}$ is a vector representing the variables present in the expression q. Taking into account the expressions from the AGCA (Aggregate Calculus[?]) the definition of an expression will be:

$$q \text{ ::- } q \cdot q | q + q | v \leftarrow q | v_1 \theta v_2 | R(\vec{y}) | \text{constant} | \text{variable}$$

We will start with the definition of $\mathsf{dom}_{\vec{x}}(R(\vec{y}))$:

$$\mathsf{dom}_{\vec{x}}(R(\vec{y})) = \left\{ \vec{c} \,|\, (\pi_{\forall i : x_i \in \vec{y}}(x_i \leftarrow c_i) \cdot R(\vec{y})) \neq 0 \right\}$$

where $\vec{x} =< x_1, x_2, x_3, \cdots, x_i, \cdots >$, $\vec{c} =< c_1, c_2, c_3, \cdots, c_i, \cdots >$. $\vec{x}$ defines the schema of $\vec{c}$, Specifically, $\vec{x}$ is the vector of names of each element of the tuple $\vec{c}$.

For the comparison operator $(v_1 \theta v_2)$, where v_1 and v_2 are variables, we can compute the domain as follows:

$$\mathsf{dom}_{\vec{x}}(v_1 \theta v_2) = \left\{ \vec{c} \,|\, \forall i : ((\pi_i(x_i \leftarrow c_i))(v_1 \theta v_2)) \neq 0 \right\}$$

The domain of a comparison is infinite. In fact using implication operator in the above definition allows us to extend the $\vec{x}$ to whatever vector we want. It is not necessary that $\vec{x}$ has the same schema as the given expression. Thus, we can evaluate $\mathsf{dom}_{\vec{x}}$ for a broader range of $\vec{x}$ and it is not delimited by the schema of the expression. In such cases the dom is infinite as the other variables can take any value. For the join operator we can write:

$$\mathsf{dom}_{\vec{x}}(q_1 \cdot q_2) = \{\vec{c}|\vec{c} \in \mathsf{dom}_{\vec{x}}(q_1) \wedge \vec{c} \in \mathsf{dom}_{\vec{x}}(q_2)\}$$

while for the union operator the domain definition is very similar:

$$\mathsf{dom}_{\vec{x}}(q_1 + q_2) = \{\vec{c}|\vec{c} \in \mathsf{dom}_{\vec{x}}(q_1) \vee \vec{c} \in \mathsf{dom}_{\vec{x}}(q_2)\}$$

$$\mathsf{dom}_{\vec{x}}(constant) = \left\{\vec{c}\right\}$$

$$\mathsf{dom}_{\vec{x}}(variable) = \left\{\vec{c}\right\}$$

Finally, we can give a formalism for expressing the domain of a variable that will participate in an assignment operation:

$$\mathsf{dom}_{\vec{x}}(v \leftarrow q_1) = \left\{\vec{c}|\vec{c} \in \mathsf{dom}_{\vec{x}}(q_1) \wedge \big(\forall i : (x_i = v) \Rightarrow (q_1 \cdot \prod_{j:x_j \neq v} (x_j = c_j) = c_i)\big)\right\}$$

Having the definitions for the domains, we will try to give some insight regarding to the notion of arity. We will start with an example relation:

$$q = R(a, b) \cdot S(b, c)$$

the schema for q will have three variables (a, b, c) The arity of a tuple will be the number of occurrences in the relation, in other words the order of multiplicity of that tuple. The arity of a tuple will increase or decrease if insertions or respectively deletions will be made to a relation. For example:

R	a	b	arity
	<1	2>	1
	<1	3>	2
	<3	4>	1

Table 1: Relation R

We need to maintain the maps for certain values as long as the arity of a tuple is greater then 0. If the arity of the tuple drops to 0 then the tuple

S	b	c	arity
	<1	1>	1
	<2	2>	2
	<3	5>	2

Table 2: Relation S

$R \cdot S$	a	b	c	arity
	<1	2	2>	2
	<1	3	5>	4
	<3	4	$*>$	0
	$<*$	1	1>	0

Table 3: Relation $R \cdot S$

will not be taken into consideration and therefore it can be eliminated from the domains of the maps.

We need to store the arities inside each map. We need a way to compute the arity of an AGCA expression. Since we substitute the subexpressions with maps, we can easily consider the maps without input variables as some relations. Also a map with input variables can be seen as a relation with a group-by clause. Input variable of a map bind some variables. Thus if we compute the map values by all different combinations of these variables, we will look up into these values and return the appropriate value according to the input variables.

2 Computing the arities of the AGCA expression

We define the function arity which will be used to compute the arity of a tuple in a certain relation. The function will be defined on the relation and the tuple for which the multiplicity order is desired to be computed.

$$arity(Relation\ \mathrm{q}, Tuple\ \mathrm{t}) = \text{multiplicity order of tuple } t \text{ in the relation } q$$

$$arity(q, t) = \pi_t(q)$$

where $\pi_t(q)$ means the projection of relation q for the tuple t. This function can be used for the computation of the arity of the expressions from the

AGCA: q ::- $q{\cdot}q \mid q{+}q \mid q\theta t \mid t \leftarrow q \mid \mathit{constant} \mid \mathit{variable}$. However constants and variables can be eliminated from the computation because relations are of interest.

$$\mathit{arity}(q_1 \cdot q_2, t) = \sum_{\{t_1\} \bowtie \{t_2\} = t} \mathit{arity}(q_1, t_1) * \mathit{arity}(q_2, t_2)$$

$$\mathit{arity}(q_1 + q_2, t) = \mathit{arity}(q_1, t) + \mathit{arity}(q_2, t), \text{ where } \mathrm{Schema}(q_1) = \mathrm{Schema}(q_2)$$

$$\mathit{arity}(v_1 \ \theta \ v_2, t =< \cdots, v_1, \cdots, v_2, \cdots >) = \begin{cases} 0, & \text{if } v_1 \theta v_2 \text{ is false} \\ 1, & \text{if } v_1 \theta v_2 \text{ is true} \end{cases}$$

$$\mathit{arity}(v \leftarrow q, t) = \begin{cases} 0, \ \text{if } \forall \vec{x} : \mathsf{dom}_{\vec{x}}(q) \text{ is empty} \\ 1, \ \text{otherwise} \end{cases}$$