# GENERALIZED JOINS

M.LACROIX
A.PIROTTE

MBLE Research Laboratory
2 avenue Van Becelaere
1170 Brussels - Belgium

In [ 1 ], Codd proposes a three-valued logic to describe the behavior of null values in the usual operations of a relational algebra.

This short note presents generalizations of the join operation which are defined in such a manner that no information is lost in the join, i.e. that the operand relations can be recovered from the join. The proposed operations introduce null values (denoted by "ω") in the join of two relations which are supposed to contain no null value. If they do contain ω, then the joins are easily modified by applying the rules given in [ 1 ].

The definition of the generalized joins are given below for the join of R(A, B1) with S(B2, C), where B1 and B2 participate in the join. Modifying the number of domains of the relations is simply a matter of notations. The following extensions will be considered as an example :

R

| A | B1 |
|---|---|
| u | m |
| u | p |
| w | n |
| x | p |
| x | r |

S

| B2 | C |
|---|---|
| n | a |
| p | b |
| q | c |

R [ B1 $\overset{+}{=}$ B2 ] S

| A | B1 | B2 | S |
|---|---|---|---|
| u | p | p | b |
| w | n | n | a |
| x | p | p | b |
| u | m | ω | ω |
| x | r | ω | ω |
| ω | ω | q | c |

The generalized equi-join R [ B1 $\overset{+}{=}$ B2 ] S is shown above for the example. The generalized θ-join of R and S can be written as :

$$R\,[\,2 \overset{+}{\theta} 1\,]\,S = R\,[\,2\theta 1\,]\,S \;\mathbf{U}\; (R-(R\,[\,2\theta 1\,]\,S)\,[\,1,2\,]) \otimes \{\langle\omega,\omega\rangle\}$$

$$\mathbf{U}\; \{\langle\omega,\omega\rangle\} \otimes (S-(R\,[\,2\theta 1\,]\,S)\,[\,3,4\,]).$$

where '⊗' denotes the cartesian product. The generalized equi-join can also be written as follows, where '*' denotes the usual natural join :

$$R\,[\,2 \pm 1\,]\,S = R\,[\,2=1\,]\,S \;\mathbf{U}\; (R*(R\,[\,2\,]-S\,[\,1\,])) \otimes \{\langle\omega,\omega\rangle\}$$

$$\mathbf{U}\; \{\langle\omega,\omega\rangle\} \otimes (S*(S\,[\,1\,]-R\,[\,2\,])).$$

A remarkable property of the generalized equi-join defined above is that, in the result, the projections on the two domains participating in the join are, in general, not equal. In practice, it is easier to attach an interesting semantics to a generalized natural join, where only one participating domain appears in the result.

This operation produces a relation whose projection on the common domain on which the join is performed is the union of the corresponding projections in the operands. On the common part of the union, the generalized natural join is the ordinary natural join.

$$R \overset{+}{*} S = R*S \cup (R*(R[2]-S[1])) \otimes \{\omega\} \cup \{\omega\} \otimes (S*(S[1]-R[2])).$$

The following relation is obtained for the example above :

R $\overset{+}{*}$ S

| A | B1=B2 | C |
|---|---|---|
| u | p | b |
| w | n | a |
| x | p | b |
| u | m | ω |
| x | r | ω |
| ω | q | c |

Remark that the generalized natural join can be seen as an ordinary natural join if R is augmented with tuples whose B1 component is in S [ B2 ] - R [ B1 ] and whose A component is ω, and S is augmented in a symmetric manner.

The generalized natural join can be modified into dissymmetric joins from which only one operand (R or S) can be obtained with no loss by projection :

$$R{+}{*}S = R*S \cup (R*(R[2]-S[1])) \otimes \{\omega\} .$$

$$R{*}{+}S = R*S \cup \{\omega\} \otimes (S*(S[1]-R[2])).$$

Obviously, $R \overset{+}{*} S = R{+}{*}S \cup R{*}{+}S$. In the example :

R+*S

| A | B1=B2 | C |
|---|---|---|
| u | p | b |
| w | n | a |
| x | p | b |
| u | m | ω |
| x | r | ω |

R*+S

| A | B1=B2 | C |
|---|---|---|
| u | p | b |
| w | n | a |
| x | p | b |
| ω | q | c |

In practice, which form to select depends on the semantics of the relations and queries. R+*S does not introduce ω in the A component, whereas R*+S does not introduce ω in the C component.

[ 1 ] CODD, E.F. Understanding relations, FDT vol. 7, N° 3-4 (23-28), 1975.