Complementary Material for Relational Conditional Set Operations

Alexis I. Aspauza Lescano[1] [0000−0001−8685−010X] and Robson L. F. Cordeiro[1] [0000−0002−6795−3004]

University of São Paulo – USP, Av. Trab. São Carlense 400, São Carlos, SP, Brazil

This is a complementary document related to the paper: "Relational Conditional Set Operations" submitted to ADBIS 2021. Here, we divide this document in two sections: (1) the algorithm to convert an infix expression to postfix, and (2) the extended relational conditional set operations which support index structures either in left or right relation.

1 Infix To Postifx Expression

We present in this section the Algorithm 1, which converts an infix expression to a postfix expression. We can see some examples of these conversions in Table 1.

Table 1: Examples of Infix and Posfix equivalences.

Infix Expression	Postfix Expression
DP.Category = SP.Category	DP.Category SP.Category =
DP.Category = SP.Category AND DP.Product = SP.Product	DP.Category SP.Category = DP.Product SP.Product = AND
DP.Category = SP.Category AND DP.Product = SP.Product AND (DP.Units ≤ SP.Units AND ¬ (DP.Price < SP.Price) OR DP.Price / 2 ≥ SP.Price)	DP.Category SP.Category = DP.Product SP.Product = AND DP.Units SP.Units ≤ DP.Price SP.Price < ¬ AND DP.Price 2 / SP.Price ≥ OR AND

Starting the algorithm, in lines 1 and 2, we need to create a stack of pending operations and a vector to contain the postfix expression to be returned. Also, in the initialization, line 3 pushes an open bracket "(" onto the stack of pending operations, as well as line 4, adds the closing bracket ")" to the end of our input expression. Lines 5 to 23 are the main loop, which will get every single token extracted from the original expression in infix notation ($infixExp$). The current token being read will be processed according to its type. If the token is an operand (any constant or reference to a column table), it will be added directly to the postfix expression vector. In case the token is an opening bracket "(", it

Algorithm 1 ToPostfix($infixExp$)

```
Input: infixExp: Expression in infix notation
Output: postfixExp: Expression in postfix notation
create stack pendingOps;
create vector postfixExp;
push "(" onto pendingOps;
add ")" to the end of infixExp;
for each token extracted from infixExp do
    if token is an operand then
        add token to postfixExp;
    else if token is "(" then
        add token to pendingOps;
    else if token is ")" then
        lastToken = pop from pendingOps;
        while lastToken != "(" do
            add lastToken to postfixExp;
            lastToken = pop from pendingOps;
        end while
    else if token is an operator then
        while top of pendingOps is an operator and it has equal or higher precedence than token do
            lastToken = pop from pendingOps;
            add lastToken to postfixExp;
        end while
        add token to pendingOps;
    end if
end for
return postfixExp;
```

will be added to the stack of pending operations. But if the token is a closing bracket ")", we will pop tokens from the pending operations and add them to the postfix expression until an opening bracket is found, and the bracket will just be discarded. As the last option, if the token is an operator of any type (logical negation, arithmetic operator, logical operator, or logical connector), we will add to the postfix expression all top operators from the pending operations while the top is an operator with higher or equal precedence than the current token; and then, add the token to the stack of pending operations. The next operators are ordered by higher to lower precedence:

1. $\neg$
2. *, / both with same precedence
3. +, - both with same precedence
4. $<, \leq, >, \geq, =, \neq$ both with same precedence
5. $\wedge$
6. $\vee$

Finally, we will return our postfix expression vector.

2 Relational Conditional Set Operations

In the paper, we assumed the index structures are always in the right relation. Here, we present an algorithm that is capable of executing the RelCond Set Operations if there are indexes either for the left or for the right relation.

Algorithm 2 RelCondSetOp(${}_c\mathsf{T}_1,\ {}_c\mathsf{T}_2,\ c,\ SetOp$)

Input: ${}_c\mathsf{T}_1$, ${}_c\mathsf{T}_2$: relational conditional sets, c: predicate, $SetOp$: $\cap_c$ or $-_c$
Output: the result ${}_c\mathsf{T}_R$ from ${}_c\mathsf{T}_1 \cap_c {}_c\mathsf{T}_2$ or from ${}_c\mathsf{T}_1 -_c {}_c\mathsf{T}_2$, according to $SetOp$

1: Create a vector of bits R with size of $|{}_c\mathsf{T}_1|$;
2: **if** $SetOp$ is $\cap_c$ **then**
3:     Initialize R with 0's;
4: **else if** $SetOp$ is $-_c$ **then**
5:     Initialize R with 1's;
6: **end if**
7: **if** ${}_c\mathsf{T}_2$ is the table with index structures **then**
8:     **for** each tuple t_i of ${}_c\mathsf{T}_1$ **do**
9:         **if** IsCondMember(${}_c\mathsf{T}_2$, t_i, c) **then**
10:             **if** $SetOp$ is $\cap_c$ **then**
11:                 set $R[i]$ as 1;
12:             **else**
13:                 set $R[i]$ as 0;
14:             **end if**
15:         **end if**
16:     **end for**
17: **else if** ${}_c\mathsf{T}_1$ is the table with index structures **then**
18:     **for** each tuple t_j in ${}_c\mathsf{T}_2$ **do**
19:         S = Index_TupleQuery(${}_c\mathsf{T}_1$, t_j, c);
20:         **if** $SetOp$ is $\cap_c$ **then**
21:             $R = R \vee S$;
22:         **else**
23:             $R = R \wedge \neg S$;
24:         **end if**
25:     **end for**
26: **end if**
27: ${}_c\mathsf{T}_R$ = get tuples for all true bits of R;
28: **return** ${}_c\mathsf{T}_R$;

Algorithm 2 implements both the conditional intersection $\cap_c$ and the conditional difference $-_c$. It receives as parameters the left relation ${}_c\mathsf{T}_1$, the right relation ${}_c\mathsf{T}_2$, the predicate c, and an indicator $SetOp \in \{\cap_c,\ -_c\}$ of the operation of interest. The result ${}_c\mathsf{T}_R$ is either ${}_c\mathsf{T}_1 \cap_c {}_c\mathsf{T}_2$ or ${}_c\mathsf{T}_1 -_c {}_c\mathsf{T}_2$, according to $SetOp$. As it is shown in Lines 1-6, the algorithm begins by creating an array of bits R to be used later to indicate the tuples of ${}_c\mathsf{T}_1$ that should be in ${}_c\mathsf{T}_R$. If $SetOp = \cap_c$, R is initialized with 0s; otherwise, it receives 1s. Now, as a first possibility as it is shown in lines 7-16, is that the right relation is the one with index structures. This is the case we presented in the paper. Here, there is a loop in lines 8-16 iterating for each tuple $t_i \in {}_c\mathsf{T}_1$. The loop updates array R only when the current tuple t_i is a conditional member of ${}_c\mathsf{T}_2$. In this case, $R[i]$ receives 1 if $SetOp = \cap_c$; otherwise, $R[i]$ is set to 0. As the second possibility, we could have the index structures not in the right relation but in the left one, as it is shown in lines 17-26. Here, there is a loop in lines 18-26 iterating for each tuple $t_j \in {}_c\mathsf{T}_2$. The first step in this loop is to compute the sub-result S as a bits vector indicating for each tuple of ${}_c\mathsf{T}_1$ if it satisfies the predicate with t_j. In this case, we will update our result R with a bit-wise operation of $R \vee S$ if $SetOp = \cap_c$; otherwise, R is set to $R \wedge S$. Finally, ${}_c\mathsf{T}_R$ is obtained as being the tuples of ${}_c\mathsf{T}_1$ that refer to each bit 1 in R.