CS345 Notes for Lecture 10/21/96

Proof L/S Algorithm Works: Only-If

Suppose $Q_1 \subseteq Q_2$. Let D be any of the canonical DB's.

- If $Q_1(D)$ contains the frozen head of Q_1, then since $Q_1 \subseteq Q_2$, so does $Q_2(D)$.
- Therefore, the L/S test will be positive for all canonical DB's.

Proof of L/S: If

Assume that for every canonical DB, the L/S test is positive. Let D be some DB, and suppose $Q_1(D)$ contains tuple t. We must show $Q_2(D)$ contains t.

- Let σ be the substitution for the variables of Q_1 that yields t.
  - ☐ I.e., for every positive subgoal G of Q_1, $\sigma(G)$ is in D, and for no negative subgoal F of Q_1 is $\sigma(F)$ in D. Also, $\sigma(H_1) = t$ (H_i is the head of Q_i).
- Partition the variables of Q_1 according to the value σ assigns them.
- This partition yields a basic canonical DB, C.
- Let τ be the 1-1 correspondence from the symbols of C to the symbols of D such that if σ maps variable X to constant a of D, and canonical DB C uses constant b for the block of the partition that contains X, then $\tau(b) = a$.
- It must be that $Q_1(C)$ contains the frozen H_1.
  - ☐ For each positive subgoal of Q_1 is mapped by σ to a member of D, and τ^{-1} maps that tuple to a member of C (the one formed from that positive subgoal).
  - ☐ And each negative subgoal of Q_1 is mapped by σ to something not in D, and therefore, $\tau^{-1} \circ \sigma$ cannot map this negative subgoal to a member of C (because

$\tau(C)$ is a subset of D).

- Now consider C', the extended canonical DB that is formed from C by taking all tuples of D formed from the symbols that are in the range of σ and applying τ^{-1} to them.
  - ☐ Thus, C' "looks like" that part of D consisting of symbols that were involved in the demonstration that t is in $Q_1(D)$.
- Since $Q_1(C)$ contains its own frozen head, we had to conduct the more extensive test where we looked at the supersets of C.
  - ☐ Evidently, these were all positive.
- Since C' was constructed from D, it must be that $Q_1(C')$ also contains Q_1's frozen head.
- Thus, $Q_2(C')$ also contains the frozen head of Q_1.
- If we apply τ to the constants involved in the demonstration that $Q_2(C')$ contains the frozen head of Q_1, we have a demonstration that $Q_2(D)$ contains t.
  - ☐ Remember that D is identical to $\tau(C')$ on the symbols that are in the range of σ.

  Thus, t is in $Q_2(D)$, completing the proof.

CQ's With Arithmetic

Suppose we allow subgoals with $<$, $\neq$, and other comparison operators.

- Negated subgoals also permitted.
- We must assume database constants can be compared.
- Technique is a generalization of the L/S algorithm, but it is due to Tony Klug (an interesting story).
- We shall work the case where $<$ is a total order; other assumptions lead to other algorithms, and we shall later give an all-purpose technique using a different approach.

Example: Consider the rules:

```
C1: p(X,Z) :- a(X,Y) & a(Y,Z) & X<Y
C2: p(A,C) :- a(A,B) & a(B,C) & A<C
```

- Both ask for paths of length 2. But Q_1 requires that the first node be numerically less than the second, while Q_2 requires that the first node be numerically less than the third.

Klug/Levy/Sagiv Test

As before, construct a family of canonical databases by considering all partitions of the variables of Q_1 (assuming we are testing $Q_1 \subseteq Q_2$).

- However, now we need to consider also the order of the values we assign to each partition.
- And if there is negation, we also need to consider extended canonical DB's; without negation, the basic canonical DB's are sufficient.

Example: To test $C_1 \subseteq C_2$ (the two CQ's of the previous example) we again need to consider the partitions of $\{X, Y, Z\}$. But now, order of the values counts too.

- The number of different basic canonical databases is 13.
  - □ For partition $\{X\}\{Y\}\{Z\}$ we have $3! = 6$ possible orders of the blocks.
  - □ For the three partitions that group two variables and leave the other separate we have 2 different orders.
  - □ For the partition that groups all three, there is one order.
- In this example, the containment test fails. We have only to find one of the 13 cases to show failure.
- For instance, consider $X = Z = 0$ and $Y = 1$. The canonical database D for this case is $\{a(0, 1), a(1, 0)\}$, and since $X < Y$, the body of C_1 is true.

- Thus, $C_1(D)$ includes $p(0, 0)$, the frozen head of C_1.
- However, no assignment of values to A, B, and C makes all three subgoals of C_2 true, when D is the database.
- Thus, $p(0, 0)$ is not in $C_2(D)$, and D is a counterexample to $C_1 \subseteq C_2$.