Expressiveness of predicate logic: H&R, p137-9

Suppose that the predicate R(x,y) expresses the fact that two nodes x and y in a graph are connected by an edge. Is it possible to write down a predicate logic formula that expresses **REACHABILITY**: the fact that two particular nodes in a directed graph G are connected by a path of directed edges? That is to say, can we find a formula φ such that φ(G,u,v) holds if and only if u and v are connected by a path in G?

In representing a general directed graph, we can introduce a predicate R(x,y) to express the fact that there is a directed edge from x to y. One way to attempt to write such a formula might then be to consider whether the following (spurious) "infinite formula" could be replaced by a semantically equivalent finite formula:

$$u = v \lor R(u,v) \lor \exists x(R(u,x) \land R(x,v)) \lor \exists x_1\, \exists x_2\, (R(u,x_1) \land R(x_1,x_2) \land R(x_2,v)) \lor \ldots$$

This is an infinite expression, so it's not a well-formed formula.

The following general theorem will be used to show that REACHABILITY is not expressible:

Theorem 2.24 (Compactness Theorem) Let Γ be a set of formulas of predicate logic. If all finite subsets of Γ are satisfiable, then so is Γ.

Proof by contradiction: Assume that Γ is not satisfiable. Then the semantic entailment $\Gamma \models \bot$ holds as there is no model in which all the formula in Γ are valid. By completeness (applied potentially in a context where the set of premises is infinite, but legitimate) this means that $\Gamma \vdash \bot$. This sequent then has a proof by natural deduction, which - being a finite piece of text - can only use a finite number of premises from Γ. Let Δ denote this finite set of premises. Then $\Delta \vdash \bot$ and from this $\Delta \models \bot$ follows from soundness.

The following theorem is a consequence of the Compactness Theorem:

Theorem 2.26: REACHABILITY is not expressible in predicate logic. That is to say: there is no predicate calculus formula with u and v as its only free variables, and R as the only predicate symbol (of arity 2) such that Φ holds in a directed graph if and only if u to v are connected by a directed path of edges in that graph.

Proof by contradiction: Suppose that there were a formula Φ to express REACHABILITY. Let Ψ_n be a formula to express the fact that there is a path of length n connecting s and t. Then Ψ_n can be defined as:

$$\exists x_1\, \exists x_2 \ldots \exists x_{n-1}\, (R(s,x_1) \land R(x_1,x_2) \land \ldots \land R(x_{n-1},t)),$$

the formula Ψ_0 can be defined as s=t, and Ψ_1 can be defined as R(s,t).

Let Γ be the infinite set of formulae:

$$\{\neg\Psi_i \mid i \geq 0\} \cup \{\Phi[s/u][t/v]\}$$

This formula cannot be satisfiable, since it expresses the non-existence of a path of any finite length from the node denoted by s to the node denoted by t, and yet - in order for the formula to be true - u and v must be connected and there is thus a finite path from s to t upon substitution of u for s and v for t.

By the Compactness Theorem, it then follows that no finite subset of the formulae in Γ can be satisfiable. But any finite subset Δ of the formulae in Γ can certainly be satisfied by choosing as a model a directed graph in which u and v are connected by a path of length greater than any excluded by any formula of the form $\neg\Psi_i$ that appears in Δ. This contradicts the assumption that there is a formula Φ to express REACHABILITY.

A note on relational DBs

One consequence of Theorem 2.26 is that the classical specification of a relational database query language (due to E.F.Codd), as based for example on the five standard relational operators, is not expressive enough to enable us to formulate a query such as "display all the *ancestors* of a given person" given a set of relational tables that record child-parent relationships (even though we can with separate queries retrieve all mothers, grandmothers and great-grand mothers etc, and formulate a query to identify ancestors going back any finite number of generations). Many DBMSs, such as Oracle, add a "connect by" feature to SQL to address this problem.
