Database Consistency: Logic-Based Approaches

Problem set #2

Wenfei Fan

due July 10, 2007

Problem 1

ECFDs are a minor extension of CFDs by incorporating inequality. An ECFD has the same form $(R(X \rightarrow Y), T_p)$ as a CFD, except that its pattern tableau T_p may contain either a constant c, an unnamed variable '_', or a "literal" $\bar{c}$ for a constant c.

Revise the match operation as follows: η_1 *matches* η_2 if either η_2 is '_', or $\eta_1 = c$ if $\eta_2 = c$, or $\eta_1 \neq c$ if $\eta_2 = \bar{c}$. Similarly, we define how a data tuple matches a pattern tuple. The semantics of ECFDs is defined along the same lines as CFDs with the revised "match" operation.

Show that the implication problem for ECFDs is coNP-complete.

Problem 2

Develop a heuristic algorithm for determining CFD implication: given a set Σ of CFDs and another CFD φ, the algorithm returns true if $\Sigma \models \varphi$. Justify the effectiveness and efficiency of your algorithm by providing either an experimental study, or a detailed analysis (e.g., certain performance guarantee for practical special cases).

Problem 3

Consider a class $\mathcal{K}$ of unary absolute XML keys of the form $(Q, @a)$, where Q is an expression in an XPath fragment $\mathcal{X}$, and $@a$ is an attribute. An XML tree satisfies $(Q, @a)$ if and only if for any two nodes v_1, v_2 reachable via Q from the root of the tree, if $v_1.@a = v_2.@a$, then $v_1 = v_2$.

The containment problem for the XPath fragment $\mathcal{X}$ is to determine, given any expressions p_1, p_2 in $\mathcal{X}$, whether or not $p_1 \subseteq p_2$; that is, for any XML tree T, the set of nodes reachable via p_1 from the root of T is a subset of the set of nodes reachable via p_2 from the root of T.

Show that the implication problem for $\mathcal{K}$ is undecidable if the containment problem for $\mathcal{X}$ is undecidable.