Tues 11am → Noon DC3346 OH Grant
Wed 3pm → 4pm

Outline

① intro by example ✓
② relational model
③ integrity constraints
④ safety and finiteness

List titles of publications written by a single author

$$\{(t) \mid \exists pid.(PUBLICATION(pid, t) \wedge \exists a \cdot WROTE(a, pid) \wedge (\forall a1 \cdot WROTE(a1, pid) \Rightarrow a1 = a))\}$$

integrity constraints

- every boss is an employee

$$\forall x, y, z \cdot EMP(x, y, z) \rightarrow \exists u, w \cdot EMP(z, u, w)$$

→ tuple-generating dependency

- every boss manages unique dept

→ equality-generating dependency

$$\forall x_1, x_2, y_1, y_2, z \cdot (EMP(x_1, y_1, z) \wedge EMP(x_2, y_2, z) \rightarrow y_1 = y_2$$

~no boss can have a boss

reform 1: $\forall x,y,z \cdot \neg EMP(x,y,z) \vee EMP(z,-,-)$ ①

$\equiv \neg \exists x,y,z \cdot \neg ①$

$\equiv \neg \exists x,y,z \cdot EMP(x,y,z) \wedge \neg EMP(z,-,-)$

- belong to a data type
- values unique among tuples in rel'n (keys)
- values in 1 rel'n must appear in another (referential integrity)
- values cannot appear simultaneously in certain relations (disjointness)

e.g. Author IDs int → type

Pub IDs are unique → keys

ref integ { Books, journals are pubs
Components of WROTE tuples must be author/pub

View is an integrity constraint of the form

$$\forall x_1, \ldots, x_k \cdot R(x_1, \ldots, x_k) \leftrightarrow \varphi$$

for R a new relation name and $x_1, \ldots, x_k$ free variables of φ

$DB \models \Sigma \iff DB, \theta \models \varphi$ for any integ. c.
$\varphi \in \Sigma$ and any valuation θ.

DB instances must be finite!

Unsafe queries:

$\{(x,y,z) \mid book(x,y,z) \vee proceedings(x,y)\}$

$\{(y) \mid \neg \exists x.\ author(x,y)\}$

$\{(x,y) \mid x = y\}$

Def^n (Domain Independent Query):

$\{(x_1, \ldots, x_k) \mid \varphi\}$ is d.indep. if

$DB_1, \theta \models \varphi \iff DB_2, \theta \models \varphi$

for any pair of instances $DB_1 = (D_1, =, R_1, \ldots, R_k)$ and $DB_2 = (D_2, =, R_1, \ldots, R_k)$ and all θ.

Theorem

Answers to DI queries contain only values that exist in $R_1, \ldots, R_k$ (the active domain).

DI + finite DB = "safe"

Theorem

Satisfiability of 1st order formulas is undecidable.

- co-r.e. in general
- r.e. for finite databases

Proof

Reduction from PCP

Theorem

DI of 1st order queries is undecidable.

Proof

φ is satisfiable iff $\{(x,y) \mid (x=y$

Definition (Range Restricted Formulas)

φ is range restricted when, for φ_i that are also range restricted, φ has the form:

$R(x_1, \ldots, x_k)$

$\varphi_1 \wedge \varphi_2$,

$\varphi_1 \wedge (x_i = x_j)$ $(\{x_i, x_j\} \cap FV(\varphi_1) \neq \emptyset)$

$\ldots$ $\varphi_1 \wedge \neg \varphi_2$ $(FV(\varphi_2) \subseteq FV(\varphi_1))$

Theorem

Range restricted $\Rightarrow$ Domain independent

Theorem

Every DI query can be written equivalently as a range-restricted query

Proof

1. Restrict every variable in φ to active domain
2. Express the active domain using a unary query over the DB instance.

data complexity in the size of the DB for fixed query