

CPSC 304 - Tutorial 6 - Solution

Datalog

Questions 1

Consider a database consisting of the relations:

$$\begin{aligned} & \text{customer}(cid, cname, rating, salary). \\ & \text{order}(cid, iid, day, qty). \\ & \text{item}(iid, iname, type, price). \end{aligned}$$

- (a) Express the following query in Datalog. “Find the details (i.e., all attributes) of the customers who have a rating higher than 6 and earn less than \$125,000.”

$$ans(I, N, R, S) \leftarrow \text{customer}(I, N, R, S), R > 6, S < 125000.$$

- (b) Express the following queries in Datalog. “Find the various item names and their types.”

$$ans(N, T) \leftarrow \text{item}(I, N, T, P).$$

- (c) What is the Datalog query equivalent to each of the following RA queries?

$$\sigma_{(rating > 5) \vee (salary > 100000)}(\text{customer}).$$
$$\sigma_{\neg(type = 'consumer electronics' \wedge price > 50)}(\text{item}).$$

The first is equivalent to

$$ans(I, N, R, S) \leftarrow \text{customer}(I, N, R, S), R > 5.$$
$$ans(I, N, R, S) \leftarrow \text{customer}(I, N, R, S), R > 100000.$$

The second is equivalent to

$$ans(I, N, T, P) \leftarrow \text{item}(I, N, T, P), T \neq 'consumer electronics'.$$
$$ans(I, N, T, P) \leftarrow \text{item}(I, N, T, P), P \leq 50.$$

- (d) Suppose in addition to the key dependencies identified by the primary keys in the given relations, we are told that every item name uniquely determines its type, i.e., that the FD $iname \rightarrow type$ holds. Suppose there are 950 distinct item names in the *item* table. Then how many tuples will be present in the result of the following Datalog query? Explain your answer.

$$ans(N, T) \leftarrow \text{item}(I, N, T, P).$$

The datalog query projects *item* on *iname* and *type*. Projection eliminates duplicates. Two tuples in the projection are duplicates of each other exactly when their *iname* values coincide, in view of the given FD. Thus, 950 distinct *iname* values lead to 950 distinct tuples in the result of the projection.

(e) Express the following query in Datalog.

“Find the type and name of those items, whose price is at most \$100 OR which were ordered by the customer with cid=123.”

$$\begin{aligned} ans(T, N) &\leftarrow item(I, N, T, P), P \leq 100. \\ ans(T, N) &\leftarrow order(123, I, _, _), item(I, N, T, _). \end{aligned}$$

(f) Express the following query in Datalog: “Find the names of customers who did not order any laptop, i.e., did not order any item of type laptop”.

$$\begin{aligned} ans(N) &\leftarrow good(C), customer(C, N, _, _). \\ good(C) &\leftarrow customer(C, _, _, _), \neg bad(C). \\ bad(C) &\leftarrow order(C, I, _, _), item(I, _, `laptop', _). \end{aligned}$$

Equivalently, we can rewrite the above Datalog query program to:

$$\begin{aligned} good(N) &\leftarrow customer(C, N, _, _), \neg bad(C). \\ bad(C) &\leftarrow order(C, I, _, _), item(I, _, `laptop', _). \end{aligned}$$

Question 2

Consider a database consisting of the relations, where the primary key of each relation is underlined.

sailors (sid, sname, rating, age)
boats (bid, bname, color)
reserved (sid, bid, date)

Write the following queries in Datalog.

(a) Find the names of sailors who have reserved at least two different boats with the same color.

$$\begin{aligned} ans(N) &\leftarrow sailors(S, N, _, _), reserved(S, B1, _), reserved(S, B2, _), \\ &B1 \neq B2, boats(B1, _, C), boats(B2, _, C). \end{aligned}$$

(b) Find the names of sailors who have reserved all red boats.

$$\begin{aligned}
 ans(N) &\leftarrow good(S), sailors(S, N, _, _). \\
 good(S) &\leftarrow sailors(S, _, _, _), \neg bad(S). \\
 bad(S) &\leftarrow sailors(S, _, _, _), boats(B, _, \`red'), \neg witness(S, B). \\
 witness(S, B) &\leftarrow reserved(S, B, _).
 \end{aligned}$$

(c) Find the name and color of boats which are reserved by all sailors rated above 7.

$$\begin{aligned}
 ans(N, C) &\leftarrow good(B), boats(B, N, C). \\
 good(B) &\leftarrow boats(B, _, _), \neg bad(B). \\
 bad(B) &\leftarrow boats(B, _, _), sailors(S, _, R, _), R > 7, \neg witness(B, S). \\
 witness(B, S) &\leftarrow reserved(S, B, _).
 \end{aligned}$$

(d) Find the name(s) of sailors with the lowest rating.

$$\begin{aligned}
 ans(N) &\leftarrow sailors(S, N, _, _), \neg bad(S). \\
 bad(S) &\leftarrow sailors(S, _, R, _), sailors(S', _, R', _), R > R'.
 \end{aligned}$$

(e) Find the name and rating of the oldest sailor(s).

$$\begin{aligned}
 ans(N, R) &\leftarrow sailors(S, N, R, _), \neg bad(S). \\
 bad(S) &\leftarrow sailors(S, _, _, A), sailors(S', _, _, A'), A < A'.
 \end{aligned}$$

(f) Find the names of sailors who have reserved every boat reserved by those with a lower rating.
E.g., if Bob is a sailor rated at 6, Pete and Rick are sailors rated below 6, then Bob must have reserved every boat reserved by Pete as well as those reserved by Rick.

$$\begin{aligned}
 ans(N) &\leftarrow sailors(S, N, _, _), \neg bad(S). \\
 bad(S) &\leftarrow sailors(S, _, R, _), sailors(S', _, R', _), R' < R, reserved(S', B, _), \neg witness(S, B). \\
 witness(S, B) &\leftarrow reserved(S, B, _).
 \end{aligned}$$