# ECS289F — Homework 2 Solutions

February 12, 2010

**Problem 1.** *Give an example of a relational calculus query $Q$, a database instance $I$, and a $D$ where $\mathsf{adom}(I) \cup \mathsf{adom}(Q) \subseteq D \subseteq \mathbb{D}$ such that $[\![Q]\!]^{I/D}$, $[\![Q]\!]^{I/\mathsf{adom}(I)\cup\mathsf{adom}(Q)}$, and $[\![Q]\!]^{I}$ are all* distinct.

**Solution.** A trivial example is the query $Q = \{x \mid x = x\}$, an empty database instance $I$, and any $D$ such that $\emptyset \subset D \subset \mathbb{D}$. Then we have $[\![Q]\!]^{I/\mathsf{adom}(I)} = \emptyset$, $[\![Q]\!]^{I/D} = D$, and $[\![Q]\!]^{I} = \mathbb{D}$. □

**Problem 2.** *Prove that the relational algebra is domain-independent (as asserted without proof in class).*

Cancelled.

**Problem 3.** *Denote by SPC ("select-project-cross product") the fragment of the relational algebra where $-$ and $\cup$ are disallowed. Prove that any SPC expression $E$ can be rewritten equivalently in* normal form

$$\pi_{i_1,\ldots,i_k}(\{c_1\} \times \cdots \times \{c_m\} \times \sigma(P_1 \times \cdots \times P_n)),$$

*where $\sigma$ is shorthand for a sequence of selection operators (of either type $\sigma_{i=j}$ or $\sigma_{i=c}$). For example, if $R$ and $S$ are binary predicates, the SPC query*

$$\pi_1(\sigma_{2=b}(R)) \times \sigma_{1=2}(\pi_{2,1}(S) \times \{d\})$$

*may be written in normal form as*

$$\pi_{2,4,3,1}(\{d\} \times \sigma_{2=b}(\sigma_{3=4}(R \times S))).$$

**Solution.** First, we push all cross products by repeatedly using the identities

$$\begin{aligned}
\pi_{i_1,\ldots,i_k}(E_1) \times E_2 &\equiv \pi_{i_1,\ldots,i_k,m+1,\ldots,m+n}(E_1 \times E_2)\\
E_1 \times \pi_{i_1,\ldots,i_k}(E_2) &\equiv \pi_{1,\ldots,m,m+1+i_1,\ldots,m+1+i_k}(E_1 \times E_2)\\
\sigma_{i=j}(E_1) \times E_2 &\equiv \sigma_{i=j}(E_1 \times E_2)\\
E_1 \times \sigma_{i=j}(E_2) &\equiv \sigma_{i+m=j+m}(E_1 \times E_2)\\
\sigma_{i=c}(E_1) \times E_2 &\equiv \sigma_{i=c}(E_1 \times E_2)\\
E_1 \times \sigma_{i=c}(E_2) &\equiv \sigma_{i+m=c}(E_1 \times E_2)
\end{aligned}$$

where in each identity, $m$ is the arity of $E_1$ and $n$ is the arity of $E_2$. (Actually, we need generalized versions of these identities that work with multi-way cross products. This is tedious but straightforward so we'll skip it.) The result is an SPC query of the form

$$op_1(\cdots op_k(E_1 \times \cdots \times E_{n+m}) \cdots)$$

where each $op_i$ is either a selection or projection, and each $E_j$ is either a predicate or a singleton constant. Second, we bring all the constant terms to the beginning of the list of cross product using the identity

$$E_1 \times E_2 \quad \equiv \quad \pi_{m+1,\ldots,m+n,1,\ldots,m}(E_2 \times E_1)$$

(again generalized to work with multi-way cross products). This may add some projections to the $op_i$'s. The result is of the form

$$op_1(\cdots op_{k'}(\{c_1\} \times \cdots \cdot \{c_m\} \times P_1 \times \cdots \times P_n) \cdots)$$

Third, we bring all projection operators to the left using

$$\begin{aligned} \sigma_{i=j}(\pi_{k_1,\ldots,k_\ell}(E)) &\quad \equiv \quad \pi_{k_1,\ldots,k_\ell}(\sigma_{k_i=k_j}(E)) \\ \sigma_{i=c}(\pi_{k_1,\ldots,k_\ell}(E)) &\quad \equiv \quad \pi_{k_1,\ldots,k_\ell}(\sigma_{k_i=c}(E)) \end{aligned}$$

and collapse them into a single projection using

$$\pi_{i_1,\ldots,i_k}(\pi_{j_1,\ldots,j_m}(E)) \quad \equiv \quad \pi_{j_{i_1},\ldots,j_{i_k}}(E)$$

The result is of the form

$$\pi_{i_1,\ldots,i_k}(\sigma(\{c_1\} \times \cdots \times \{c_m\} \times P_1 \times \cdots \times P_n))$$

Finally, we focus on the selection operators. For each selection-1 operator $\sigma_{i=j}$, if $i \leq m$ (or $j \leq m$), it refers to a constant singleton so we replace it with $\sigma_{j=c_i}$ ($\sigma_{i=c_j}$). Then, for each selection-2 operator $\sigma_{i=c}$ where $i \leq m$, if $c_i \neq c$, then the query is unsatisfiable so we can just stop and output "unsatisifiable"[1] and stop; otherwise, we just remove it. At this point the indices for all selection operators are $> m$. We finish by pushing them back inside the cross product using the identities from the very first step. The result is of the required normal form

$$\pi_{i_1,\ldots,i_k}(\{c_1\} \times \cdots \times \{c_m\} \times \sigma(P_1 \times \cdots \times P_n)).$$

$\square$

[1] i.e., output an arbitrary unsatisfiable SPC query in normal form of the appropriate output arity, such as $\pi_{1,\ldots,1}(\{c\} \times \sigma_{1=c_1}(\sigma_{1=c_2}(R)))$.