COMP207 Lab Exercises
Tutorial 6 (Week 8)

1 Week 8 Lab Exercises

If you have not worked through all of the exercises (except, maybe, the programming exercise), you should use the time to continue working on these exercises. Model solutions have not been provided, however, it is worthwhile to try to solve these on your own. If you get stuck, you could consult one of the demonstrators to see how to proceed.

2 Relational Algebra Revisited

In Lecture 12–16 we talked about query processing and saw that Relational Algebra plays an important role in this task. To gain a deeper understanding of Relational Algebra, it is necessary to provide more precise definition of some of the concepts of the relational model (tuples, relations) and of the operators of Relational Algebra.

Let us start by giving more precise definitions of tuples and relations. You may find it useful to compare these formal definitions with the informal presentation of these concepts in the lectures as well as in COMP102/CSE103.

2.1 Tuples and Relations

A *relation schema* has the form $R(A_1,A_2,\ldots,A_n)$, where R is a relation name, $n \geq 0$, and $A_1,A_2,\ldots,A_n$ are distinct attribute names. Thus, a relation schema provides us with the name of a relation as well as with its attributes. For instance, $\mathsf{Student}(\mathsf{student\ id},\mathsf{name},\mathsf{email\ address})$ and $\mathsf{Module}(\mathsf{module\ code},\mathsf{title},\mathsf{coordinator},\mathsf{year})$ are relation schemas.

A *tuple* over a relation schema $R(A_1,A_2,\ldots,A_n)$ is a mapping t with domain $\{A_1,A_2,\ldots,A_n\}$ that assigns to each attribute A_i, for $i \in \{1,\ldots,n\}$, a value $t(A_i)$. The follwing are examples of tuples over the relation schema $\mathsf{Module}(\mathsf{module\ code},\mathsf{title},\mathsf{coordinator},\mathsf{year})$:

- the mapping t_1 with domain $\{\mathsf{module\_code},\mathsf{title},\mathsf{coordinator},\mathsf{year}\}$ and
  - $t_1(\mathsf{module\_code}) =$ COMP105,
  - $t_1(\mathsf{title}) =$ Programming Language Paradigms,
  - $t_1(\mathsf{coordinator}) =$ J. Fearnley, and
  - $t_1(\mathsf{year}) = 1$.
- the mapping t_2 with domain $\{\mathsf{module\_code},\mathsf{title},\mathsf{coordinator},\mathsf{year}\}$ and
  - $t_2(\mathsf{module\_code}) =$ COMP201,
  - $t_2(\mathsf{title}) =$ Software Engineering I,

- $t_2(\mathsf{coordinator}) = \text{S. Coope}$, and
- $t_2(\mathsf{year}) = 2$.

A *relation* over a relation schema $R(A_1, A_2, \dots, A_n)$ is a set of tuples over $R(A_1, A_2, \dots, A_n)$. For instance, the set $\{t_1, t_2\}$ that consists of the two tuples t_1, t_2 from above is a relation over $\mathsf{Module}(\mathsf{module\_code}, \mathsf{title}, \mathsf{coordinator}, \mathsf{year})$. Using the notation from the lectures, we would represent this relation as:

Module

module_code	title	coordinator	year
COMP105	Programming Language Paradigms	J. Fearnley	1
COMP201	Software Engineering I	S. Coope	2

Note that t_1 corresponds to the first row of this table, whereas t_2 corresponds to the second row.

We often do not distinguish between a relation name and the corresponding relation. So, for example, we will write $\mathsf{Module} = \{t_1, t_2\}$ for the above relation, although strictly speaking Module is a relation name. It will always be clear from the context if we mean a relation or a relation name.

Since the set representation of a relation r is: $r = \{t_1, t_2\}$ the set representation for this example is: Module = {(COMP105, Programming Language Paradigms, J. Fearnely, 1), (COMP201, Software Engineering I, S. Coope, 2)}

Exercise 1. Represent the following table using the concepts introduced above:

Student

student_id	name	email_address
COMP105	Programming Language Paradigms	J. Fearnley
COMP201	Software Engineering I	S. Coope

(a) Write down the tuple corresponding to the first row. What is the domain of this tuple. Which values are assigned to each attribute?

(b) Do the same for the second row.

(c) Write down the entire relation as above, in set representation.

Exercise 2. Represent the following table using the concepts introduced above:

Films

film_id	title	year	rating
10234	King Kong	1933	7.9
67623	King Kong	1976	5.8
83943	Footloose	1984	6.5

(a) Write down the corresponding relation schema.

(b) Write down the tuple corresponding to the first row. What is the domain of this tuple. Which values are assigned to each attribute?

(c) Do the same for the second and third row.

(d) Write down the entire relation in set representation.

2.2 Relational Algebra

We are now ready to give definitions of some of the operators of the Relational Algebra. In the following, let $R(A_1,\dots,A_m)$ and $S(B_1,\dots,B_n)$ be two relation schemas. Then,

- $\sigma_{\mathsf{condition}}(R) = \{t \in R \mid t \text{ satisfies } \mathsf{condition}\}$. The condition can be an arbitrary condition on the attributes of R, such as '$A_1 = a$ AND $A_3 = A_7$' (see the lecture notes for COMP102/CSE103, or for lecture 12).
- $\pi_{A_{i_1},\dots,A_{i_k}}(R) = \{t[A_{i_1},\dots,A_{i_k}] \mid t \in R\}$, where $t[A_{i_1},\dots,A_{i_k}]$ is the mapping with domain $\{A_{i_1},\dots,A_{i_k}\}$ that assigns to each attribute A_{i_j} the value $t(A_{i_j})$.
- $R \bowtie S = \{r \bowtie s \mid r \in R, s \in S, \text{ and } r(C) = s(C) \text{ for all } C \in \{A_1,\dots,A_m\} \cap \{B_1,\dots,B_n\}\}$, where $t \bowtie s$ is the mapping with domain $\{A_1,\dots,A_m,B_1,\dots,B_n\}$ that assigns to each attribute A_i the value $r(A_i)$ and to each attribute B_i the value $s(B_i)$.
- If R and S have no attributes in common, then $R \times S = R \bowtie S$.
- If R and S have the same attributes, then $R \cup S = \{t \mid t \in R \text{ or } t \in S\}$.
- If R and S have the same attributes, then $R \cap S = \{t \mid t \in R \text{ and } t \in S\}$.
- If R and S have the same attributes, then $R - S = \{t \mid t \in R \text{ and } t \notin S\}$.

It is now possible to prove that equations involving Relational Algebra expressions.

Example 1. Let us verify the following equality, where A is an attribute of R and S:

$$\sigma_{A=a}(R \bowtie S) = \sigma_{A=a}(R) \bowtie \sigma_{A=a}(S)$$

As you know from year 1, there are two things we have to show:

(a) $\sigma_{A=a}(R \bowtie S) \subseteq \sigma_{A=a}(R) \bowtie \sigma_{A=a}(S)$

(b) $\sigma_{A=a}(R) \bowtie \sigma_{A=a}(S) \subseteq \sigma_{A=a}(R \bowtie S)$

To prove (a), let t be a tuple in $\sigma_{A=a}(R \bowtie S)$. Then, $t \in R \bowtie S$ and $t(A) = a$. Observe that $t \in R \bowtie S$ implies that there exist tuples $r \in R$ and $s \in S$ such that r and s agree on all common attributes of R and S and $t = r \bowtie s$. Also, we have $r(A) = t(A) = a$ and $s(A) = t(A) = a$. Therefore, $r \in \sigma_{A=a}(R)$ and $s \in \sigma_{A=a}(S)$ and r and s agree on all common attributes of R and S, which implies $t = r \bowtie s \in \sigma_{A=a}(R) \bowtie \sigma_{A=a}(S)$.

To prove (b), let $t \in \sigma_{A=a}(R) \bowtie \sigma_{A=a}(S)$. Then there are tuples $r \in \sigma_{A=a}(R)$ and $s \in \sigma_{A=a}(S)$ such that r and s agree on all common attributes and $t = r \bowtie s$. Note that $r \in R$ and $s \in S$, which implies $t \in R \bowtie S$. Now, since $r \in \sigma_{A=a}(R)$ and $s \in \sigma_{A=a}(S)$, we know that $t(A) = r(A) = s(A) = a$. Therefore, $t \in \sigma_{A=a}(R \bowtie S)$.

Exercise 3. Assume two relations R and S such that attribute B occurs only in R, and attribute C occurs only in S. Which of the following equations are true, which are false? Try to prove your answers.

(a) $\sigma_{B=b \text{ AND } C=c}(R \bowtie S) = \sigma_{B=b}(R) \bowtie \sigma_{C=c}(S)$

(b) $\pi_{B,C}(R \bowtie S) = \pi_B(R) \bowtie \pi_C(S)$