SOLUTION NOTES

Databases 2002 Paper 6 Question 8 (GMB)

1. [Bookwork]. In the lecture notes, the core relational algebra was defined to consist of the following five operator (classes):

   (a) The set theoretic operators: union, intersection and difference.

   (b) Selection, $\sigma_{P(\vec{A})}$

   (c) Projection, $\pi_{\vec{A}}$.

   (d) Products and Joins: Cartesian Product $\times$, and natural joins $\bowtie$.

   (e) Renaming: $\rho_{B/A}$.

   Perfect answers will provide the typing rules for these operators, and mention the concept of **union compatibility**.

2. (a) **Full Outer Join.** We extend the core algebra to allow attribute values to be unknown/inapplicable. We use a special value NULL for this purpose. Full Outer Join is an extension of the join operator so as not to lose any dangling tuples. Since the tuples in the resulting relation have all the attributes of both relations, we pad out the dangling tuples with NULLS. There are other variations, including left and right outer joins, and theta outer joins. For example, consider the tables:

   R =

A	B
1	2
3	4

   S =

B	C
4	5
6	7

   R$\overset{\circ}{\bowtie}$ S =

A	B	C
3	4	5
1	2	NULL
NULL	6	7

   (b) **Grouping and Aggregate Operator** We employ a collection of fairly standard aggregation operators: Sum, Average, Count, Min and Max.

   The Grouping Operator $\gamma_L(R)$, where L is a list of elements that are either:

   - Individual attributes, or
   - Of the form $\Theta(A)$ where Θ is one of the aggregation operators above, and A the attribute to which its applied.

   It is computed by

   i. Group R according to the grouping attributes in L
   ii. Within each group compute $\Theta(A)$,
   iii. Result is the relation whose columns consist of one tuple for each group. The components of that tuple are the values associated with each element of L for that group.

   Take the expression $\gamma_{\text{Beer,AVG(Price)}}(R)$

R =

Bar	Beer	Price
Eagle	Abbot	2.30
Eagle	Fosters	2.50
Castle	Abbot	2.40
Castle	1664	2.60
Boathouse	Fosters	2.50

The first stage yields:

Bar	Beer	Price
Eagle	Abbot	2.30
Castle	Abbot	2.40
Eagle	Fosters	2.50
Boathouse	Fosters	2.50
Castle	1664	2.60

Computing the average then yields the relation:

Beer	Price
Abbot	2.35
Fosters	2.50
1664	2.60

3. (a) The query would be compiled to the following relational algebra term.

$$\pi_A(\sigma_{X.A=Y.A \text{ or } X.A=Z=A}(X \times Y \times Z)$$

(b) The SQL query (and the algebra) does **not** produce the result expected by the user. The problem is that SQL forms a Cartesian product of all three relations. If, for example, Z is the empty set, then the user would expect her query to yield $X \cap Y$, but the Cartesian product ensures that the SQL query will yield the empty set!