2008 Paper 6 Question 8

Databases

(a) Define the notion of a *safe query* in the relational calculus. [2 marks]

(b) Suppose that we have schemas $R(A,\ B)$ and $S(B,\ C)$, and that the number of tuples in R is r and the number of tuples in S is s. Suppose that both R and S are not empty, and that neither contains duplicates.

For each of the following relational algebra queries, state in terms of r and s the *minimum possible* and *maximum possible* number of tuples in the result.

(i) $\sigma_p(R \times S)$ [2 marks]

(ii) $\pi_{A,\ C}(R \times S)$ [2 marks]

(iii) $\pi_B(R) - (\pi_B(R) - \pi_B(S))$ [2 marks]

(iv) $R \overset{\circ}{\bowtie}_L S$ (left outerjoin) [2 marks]

(v) $R \overset{\circ}{\bowtie} S$ (full outerjoin) [2 marks]

(c) Again, suppose that we have schemas $R(A,\ B)$ and $S(B,\ C)$. Make no assumptions about functional dependencies. Let b be some value from domain B. Consider the following relational algebra queries.

1. $\pi_{A,C}(R \bowtie \sigma_{B=b}(S))$
2. $\pi_A(\sigma_{B=b}(R)) \times \pi_C(\sigma_{B=b}(S))$
3. $\pi_{A,C}(\pi_A(R) \times \sigma_{B=b}(S))$

Two of these queries always return the same result, while one may not. Which one is different? Give a simple database instance in which this query returns a different result.

[8 marks]