Database and Information Systems

Solutions to Homework 1

Due on September 27, 2005

The first two problems concern the Penn Ebay (PBAY) System, which is represented by the following schema:

- Sellers(sellerID:int, rating:char, email:string)
- Buyers(buyerID:int, email:string, address:string)
- Stock(itemID:int, sellerID:int, price:float, quantity:int, endingTime:int)
- Purchase(itemID:int, buyerID:int, sellerID:int, price:float, purchaseQuantity:int, bidTime:int)

Problem 1 [60 points]: Express the following queries in (a) the relational algebra, (b) the tuple relational calculus, and (c) the domain relational calculus:

Note: in problems where wording proved unclear, answers correct with respect to some reasonable interpretation of the problem were accepted.

1. Find the IDs of items with startBid price < $10

   RA: $\pi_{itemID}(\sigma_{startBid<10}(Stock))$
   TRC: $\{P \mid \exists S \in Stock(S.startBid < 10 \land P.itemID = S.itemID)\}$
   DRC: $\{ \langle i > \mid \exists s, b, q, e(\langle i, s, b, q, e \rangle \in Stock \land b < 10) \}$

2. Find the emails of sellers with rating 'A' who have items in stock with startBid price < $10$.

   RA: $\pi_{email}(\sigma_{rating='A'}(Sellers) \bowtie \sigma_{sellerID}(\sigma_{startBid<10}(Stock)))$
   TRC: $\{P \mid \exists S \in Sellers, \exists T \in Stock (S.rating = "A" \land S.sellerID = T.sellerID \land T.startBid < 10 \land P.email = S.email)\}$
   DRC: $\{ \langle m > \mid \exists s, r(\langle s, r, m \rangle \in Sellers \land (r = "A" \land (\exists i, b, q, e(\langle i, s, b, q, e \rangle \in Stock \land b < 10)))) \}$

3. Find the IDs of buyers who purchased items for over 10% beyond the startBid price of the items they bought.

   RA: $\pi_{buyerID}(\pi_{itemID,sellID}(Stock) \bowtie \pi_{itemID,buyID,price}(Purchase))$
   TRC: $\{P \mid \exists T \in Stock, \exists U \in Purchase (T.itemID = U.itemID \land U.price > 1.1 \ast T.startBid \land U.buyerID = P.buyerID)\}$
   DRC: $\{ \langle b > \mid \exists i, p, d, \exists s, q(\langle i, s, d, q, e \rangle \in Stock) \land \exists s, m(\langle i, b, s, p, u, m \rangle \in Purchase) \land p > 1.1 \ast d \}$

4. Find the IDs of buyers who purchase items with type "book" and with a bidTime within 5 minutes of the endingTime. (clarify: endingTime - bidTime will return minutes)

   RA: $\pi_{buyerID}(\pi_{itemID}(\sigma_{type='book'}(Items)) \bowtie \pi_{sellerID,stock,endingTime}(Stock) \bowtie \sigma_{endingTime-bidTime<5}(\pi_{sellerID,stock,buyID,bidTime}(Purchase))$
   TRC: $\{P \mid \exists I \in Items, \exists T \in Stock, \exists U \in Purchase (I.itemID = T.itemID \land T.itemID = U.itemID \land T.sellerID = U.sellerID \land (U.endingTime - T.bidTime < 5) \land I.type = "book" \land U.buyerID = P.buyerID)\}$
   DRC: $\{ \langle b > \mid \exists (\langle i, t \rangle \in Items \land t = "book") \land (\exists d, e(\langle i, t, d, q, e \rangle \in Stock) \land \exists p, u(\langle i, s, b, p, u, m \rangle \in Purchase \land (e - m) < 5)) \}$
Problem 2 [30 points]: State in English what the following queries compute:

1. $\pi_{\text{buyerID}}(\sigma_{\text{purchaseQuantity} = 2}(\text{Purchase}) \bowtie \pi_{\text{sellerID}}(\sigma_{\text{rating} = 'A'}(\text{Sellers})))$

   IDs of buyers who have bought 2 of the same items from a seller with rating “A”.

2. $\pi_{\text{email}}((\pi_{\text{sellerID}}(\sigma_{\text{quantity} = 4 \land \text{startBid} < 10}(\text{Stock})) \bowtie \pi_{\text{sellerID}}(\sigma_{\text{price} > 20}(\text{Purchase}))) \bowtie \text{Sellers})$

   Emails of sellers who have 4 of the same stock items, all with start bid price < 10, and who have sold at least 1 item with price > 20.

3. $\pi_{\text{email}}(\sigma_{\text{address} = 'philadelphia'}(\text{ Buyers}) \bowtie \pi_{\text{buyid}}(\sigma_{\text{price} < \text{startBid}}(\sigma_{\text{type} = 'book' \land \text{purchaseQuantity} = 2}(\text{Items \bowtie \text{Purchase}) \bowtie \text{Stock})))$

   Emails of the buyers living in Philadelphia who have bought 2 same books with price less than the start bid price.

4. $\pi_{\text{rating}}(\pi_{s1}(\sigma_{i1 \neq i2 \land a1 = 2}(\rho_{\text{itemID} \rightarrow i1, \text{sellerID} \rightarrow s1}(\text{Stock}) \bowtie \rho_{\text{itemID} \rightarrow i2, \text{sellerID} \rightarrow s2}(\sigma_{\text{quantity} \geq 3}(\text{Stock})))) \bowtie \text{Sellers})$

   Ratings of sellers with 2 different items in stock where the seller stocks at least 3 of one of those items.

Problem 3 [10 points]: Explain the principles of data independence, as argued for by Codd, and their significance.

The cornerstone of Codd’s ideas of data independence is the separation of the data to be processed from the means of processing it, in order to enable greater flexibility and extensibility. In essence, the logical data model is a representation of the data independent of its order, indices, and access methods or paths. These concepts are the ones upon which today’s database field is built.