CS430/630 – Homework 1

Released Feb 03, Due Feb 17

50 points (5/100 of final grade)

Instructions: The homework is **due at 19:00 EST on Wed Feb 17**. The submission must be in a **SINGLE PDF FILE** called **HW1.pdf**. The homework must be **typeset** (handwritten submissions will **NOT** be accepted). The email **MUST BE SENT FROM YOUR @UMB ACCOUNT** and must have subject formatted **EXACTLY AS BELOW**, and there must be a single attachment (HW1.pdf):

[CS430] HW1   - for CS430 students, and

[CS630] HW1   - for CS630 students.

The timestamp will be considered according to the UMB email server. Submissions received (i.e., tagged) later than 19:00 EST on Feb 17 **WILL NOT BE GRADED**. You are only allowed a **SINGLE** submission (i.e., you cannot send multiple emails – only the **first received** will be graded).

Email submissions must be sent to Nicholas J Pankewytch at <N.Pankewytch001@umb.edu>

For CS430, all questions have equal weight. For CS630, all questions (a-f) have the same weight (i.e., (25-3)/6).

Question 1 (25 points)

You are given the following relational schema.

```
Books(bid:integer, bname:string, author:string, year:integer, price:integer)
Orders(cid:integer, bid:integer, quantity:integer)
Customers(cid:integer, cname:string, zipcode:string)
```
The meaning of attributes is as follows:

`bid`: book unique identifier, `bname`: book name; author, year and price (per unit) are self-explanatory

`quantity`: number of books purchased with an order

`cid`: unique customer identifier, `cname`: customer name, `zipcode`: customer address zipcode.

Write **relational algebra** expressions for the following queries:

(a) List the publication year of books ordered from zipcode 02125.
(b) Find the zipcodes of customers who bought in a single order at least 10 copies of a book that costs more than $100.
(c) Find the names of customers who bought **only** books published in or after 1990 (implies they bought at least a book).
(d) Find the names of customers who bought at least a copy of a book by author 'Edgar Codd', or at least 10 copies of a book called 'Databases'.
(e) Find the names of customers who placed an order with a total payment of more than $1,000.
(f) Find the author(s) of the book(s) for which there is a single order placed.
(g) [630 only] Find the names of books ordered by those customers who are the only ones in their particular zipcode who placed an order (i.e., there is no other customer with the same zipcode who placed an order). (3p)

Question 2 (25 points)

You are given the following relational schema.

```
Passengers(pid:integer, pname:string, age:integer, city:string)
Flights(fid:integer, from:string, to:string, miles:integer, aircraft:string)
Tickets(pid:integer, fid:integer, price:integer)
```

The database contains passenger and flight information for an airline in a single day. The description is as follows: passengers are uniquely identified by `pid` and they have as attributes name, age and city of residence. Flights are uniquely identified by `fid`. Each flight has an origin and destination airport identified by `from` and `to` respectively (airport codes are 3-letter strings, and are unique – no two airports have the same code - you do not need to enforce that constraint, just assume it is respected), flight distance given by `miles`, and `aircraft` type, e.g., B737, B777, B787, A320, etc. The `Tickets` relation provides information of flights that passengers took, and also the `price` that was paid for that flight **segment** (assume that each ticket corresponds to a segment; for people with connecting flights, they will have separate tickets, billed independently).

Write **relational algebra** expressions for the following queries given the schema above:

(a) Find the names of passengers who traveled on a B787 aircraft and paid at most $800 for that ticket.

(b) Find the ages of passengers who traveled **only** on flights going in or out of BOS (assumes that they were on at least one flight).

(c) List the prices of all tickets for flights on aircrafts of type B777.

(d) Find the residence city of passengers who traveled on a flight longer than 500 miles, or who paid at most $500 for a ticket.

(e) Find the origin of flights that had **both** passengers who reside in Boston and passengers who reside in Chicago.

(f) Find the names of passengers who had connecting flights, defined as **two** flights where the origin of the second flight is the same as the destination of the first. Note that, return flights do not qualify as connecting flights (e.g., BOS to SFO and SFO to BOS is not a connecting flight, whereas BOS to JFK and JFK to LAX is a connecting flight).

(g) [630 only] Find the names of "suspicious" passengers, defined as passengers who took a trip consisting of three segments (say airport A -> B -> C -> D) even though there are direct flights connecting the source and final destination (A -> D). (3p)