

CSE 562: Homework #1 (due 02/17/11)

You are given the following relational schema:

MOVIE(TITLE,DIR, YEAR, BUDGET)
CAST(TITLE,ACTOR).

Keys are underlined. The attribute TITLE in CAST is a foreign key referencing MOVIE(TITLE).

Problem 1 (24 pts)

Write the following queries in relational algebra and SQL:

- **1.1:** Find all the directors of the movies starring Harrison Ford;
- **1.2:** Find all the actors that starred in “Star Wars” but not in “Return of the Jedi”;
- **1.3:** Find all the actors that starred in all movies starring Harrison Ford.

Write the following queries in SQL:

- **1.4:** For every actor, list the actor and the number of movies he/she made under the direction of Steven Spielberg;
- **1.5:** Find all the actors with the highest number of movies whose budget was over the average movie budget.

Problem 2 (8 pts)

You are given a relation schema CONNECTION(ORIGIN,DESTINATION,AIRLINE) representing airline connections. Write the following queries in SQL, possibly using recursion:

- **2.1:** For every city, find the number of cities connected to it using a single airline with at most one stopover. The connections may be indirect.
- **2.2:** Find all the cities connected to Buffalo using arbitrary airlines with an arbitrarily large number of stopovers. The stopovers should not include Boston.

Note: Please bear in mind that Oracle does not support recursion. If you want to test your recursive queries, use PostgreSQL.

Problem 3 (12 pts)

Consider the following SQL queries:

- **3.1:**

```
SELECT c2.ACTOR
FROM MOVIE m, CAST c1, CAST c2
WHERE m.TITLE=c1.TITLE
      AND m.TITLE=c2.TITLE
      AND c1.ACTOR='Harrison Ford'
      AND m.DIR='Steven Spielberg'
```

- **3.2:**

```
SELECT DIR, MIN(YEAR)
FROM MOVIE
GROUP BY DIR
```

For both queries:

- Explain what the query is doing.
- Translate the query to relational algebra.

Problem 4 (6 pts)

You are given a relation $EMP(NAME, DEPT, SALARY)$. Explain what each query below is doing.

Two queries are *equivalent* if they return the same answer for every database instance. Determine if the queries below are equivalent. If the queries are equivalent, argue why this is the case. If the queries are not equivalent, provide a database instance for which they return different results.

Query 4.1:

```
SELECT e1.NAME From EMP e1
WHERE NOT EXISTS
(SELECT * FROM EMP e2
WHERE e1.DEPT=e2.DEPT AND e1.SALARY>e2.SALARY)
```

and 4.2:

```
SELECT e1.NAME FROM EMP e1
WHERE NOT EXISTS
(SELECT * FROM EMP e2, EMP e3
WHERE e1.DEPT=e2.DEPT AND e2.DEPT=e3.DEPT AND e1.SALARY>e3.SALARY)
```

Problem 5 extra credit(10 pts)

You are given a relation R with n numeric columns. Write a query in SQL that returns the tuples having the *maximum number of occurrences of the value 1*. The query should have size at most quadratic in n .