Name – Sithalechumi Narayanan
Date – 30th September 2015
BINF 650 – Fall 2015
Homework – 4

This assignment uses movie database.

1. In a single command return the aid, firstname and lastname of all of the actors who have two names in their first name. Exclude those that have just an initial. For example, "Mary Beth" is acceptable but "C. Everett" is not. Remember that it is possible that there are unseen characters at the ends of some names.

```
binf:~ snaray11$ mysql -h binf.gmu.edu -D binf650 -u snaray11 -p
Enter password:
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 13348
Server version: 5.1.73 Source distribution

Type 'help;' or '\h' for help. Type '\c' to clear the current input
statement.
```

```
mysql> SHOW DATABASES;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| binf650            |
| mysql              |
| snaray11           |
+--------------------+
4 rows in set (0.00 sec)
```

```
mysql> USE binf650;
Database changed
mysql> SHOW TABLES;
+------------------+
| Tables_in_binf650 |
+------------------+
| actors           |
| author           |
| genome           |
| isin             |
| movies           |
| protein          |
| published        |
| splices          |
+------------------+
8 rows in set (0.01 sec)
```

```
mysql> DESC actors;
+-----------+-------------+------+-----+---------+----------------+
| Field     | Type        | Null | Key | Default | Extra          |
+-----------+-------------+------+-----+---------+----------------+
| aid       | int(11)     | NO   | PRI | NULL    | auto_increment |
| firstname | varchar(25) | YES  |     | NULL    |                |
| lastname  | varchar(25) | YES  |     | NULL    |                |
+-----------+-------------+------+-----+---------+----------------+
3 rows in set (0.01 sec)
```

```
mysql> SELECT aid, firstname, lastname FROM actors WHERE
rtrim(firstname) like "% %" AND firstname NOT LIKE "% _.%" AND
firstname NOT LIKE "_.%";
+-----+---------------+----------+
| aid | firstname     | lastname |
+-----+---------------+----------+
|   6 | James (Jimmy) | Stewart  |
|  78 | Tim Blake     | Nelson   |
|  93 | Penelope Ann  | Miller   |
| 102 | David Hyde    | Pierce   |
| 153 | Tommy Lee     | Jones    |
| 203 | Lesley Ann    | Warren   |
| 218 | Joe Don       | Baker    |
| 229 | Billy Bob     | Thornton |
| 254 | David Ogden   | Stiers   |
| 277 | Philip Seymour | Hoffman |
| 292 | Jamie Lee     | Curtis   |
| 300 | Billy Dee     | Williams |
| 307 | Sarah Jessica | Parker   |
| 316 | Helen Bonham  | Carter   |
| 325 | James Earl    | Jones    |
| 561 | Jennifer Jason | Leigh   |
| 605 | Weird Al      | Yankovic |
| 616 | Niel Patrick  | Harris   |
| 628 | Edward James  | Olmos    |
| 640 | Thomas Haden  | Church   |
| 645 | Lee Van       | Cleef    |
| 650 | Marcia Gay    | Harden   |
| 661 | Neil Patrick  | Harris   |
| 686 | Mary Kay      | Place    |
+-----+---------------+----------+
24 rows in set (0.00 sec)
```

I used rtrim to remove the trailing ends of a string so I don't end up getting first names that just have only one name in it.

2. In a single command return the actor id and lastname of the 10 actors that has the letter "er" the farthest away from the beginning of their lastname. For example, the name "Unser" has 3 letters before "er" and "Alabaster" has 7 letters before the "er". Some names like "Ferrer has 2 "er" combinations. Use only the last ones. This list must be sorted according to the distance from the beginning of the last name to the last occurrence of "er" with the largest distance being the first entry in the sorted list.

```
mysql> SELECT aid, lastname, LENGTH(lastname) - LOCATE('re',
REVERSE(lastname)) AS lastoccurrence FROM actor
s WHERE lastname LIKE "%er%" ORDER BY lastoccurrence DESC LIMIT 10;
+-----+----------------+----------------+
| aid | lastname       | lastoccurrence |
+-----+----------------+----------------+
|  92 | Schwarzenegger |             13 |
|  34 | Ratzenberger   |             11 |
|  98 | Lancaster      |              8 |
| 116 | Zellweger      |              8 |
| 572 | Klemperer      |              8 |
| 565 | Schneider      |              8 |
| 552 | Straithern     |              8 |
| 634 | Schreiber      |              8 |
| 482 | Guttenberg     |              8 |
| 299 | Basinger       |              7 |
+-----+----------------+----------------+
10 rows in set (0.01 sec)
```

I am locating “re” instead of “er” because I want to check the name for the farthest “er”. And so I start in the reverse order and check for the occurrence of “er” and then subtract it by the length of the string.

3. Return the title of the movie that has the most languages (which is defined as the longest string for the language variable).

```
mysql> SELECT title, language FROM movies ORDER BY LENGTH(language)
DESC LIMIT 1;
+--------+-------------------------------------------------------------------------------
-+
| title  | language
|
+--------+-------------------------------------------------------------------------------
-+
| Munich | English  | German  | French  | Hebrew  | Arabic  | Italian  | Greek  | Russian
|
+--------+-------------------------------------------------------------------------------
-+
1 row in set (0.01 sec)
```

4. Have you heard of the Kevin Bacon effect? The idea is that every actor is only a few handshakes away from Kevin Bacon. In this case, a handshake is replaced by a movie that has common actors. For example, Kevin Bacon and Kathy Bates are only two handshakes apart. Kevin Bacon and Bill Paxton were both in the movie Apollo 13 thus they shook hands. Bill Paxton and Kathy Bates were both in Titanic and so they shook hands. Bacon shakes Paxton and Paxton shakes Bates. Thus, Bacon and Bates are connected by two handshakes. Your job is to find the two handshake path between Kevin Bacon and Bill Murray. This will take multiple queries – please show all your work.

```
mysql> SELECT aid FROM actors WHERE firstname LIKE 'Kevin' AND lastname
LIKE 'Bacon';
```

```
+-----+
| aid |
+-----+
| 143 |
+-----+
1 row in set (0.00 sec)
```

```
mysql> SELECT aid FROM actors WHERE firstname LIKE 'Bill' AND lastname
LIKE 'Murray';
+-----+
| aid |
+-----+
| 245 |
+-----+
1 row in set (0.00 sec)
```

Using the actor id for Kevin Bacon and Bill Murray, I found the list of actors who acted with them in their movies respectively.

The following list if for Kevin Bacon whose aid is 143.

```
mysql> SELECT DISTINCT actors.aid, actors.firstname, actors.lastname
FROM actors, isin WHERE actors.aid = isin.actor AND isin.mid IN (SELECT
mid FROM isin WHERE actor = 143)ORDER BY actors.aid;
+-----+------------+-----------+
| aid | firstname  | lastname  |
+-----+------------+-----------+
|   9 | Tom        | Hanks     |
|  22 | John       | Candy     |
| 111 | Tim        | Robbins   |
| 122 | Steve      | Martin    |
| 128 | Ed         | Harris    |
| 139 | Sean       | Penn      |
| 142 | Forest     | Whitaker  |
| 143 | Kevin      | Bacon     |
| 144 | Brendan    | Fraser    |
| 145 | Andy       | Garcia    |
| 194 | Fred       | Ward      |
| 198 | Bill       | Paxton    |
| 213 | Kathleen   | Quinlan   |
| 252 | Laura      | Linney    |
| 555 | Laurence   | Fishburne |
| 650 | Marcia Gay | Harden    |
+-----+------------+-----------+
16 rows in set (1.79 sec)
```

The following list if for Bill Murray whose aid is 245.

```
mysql> SELECT DISTINCT actors.aid, actors.firstname, actors.lastname
FROM actors, isin WHERE actors.aid = isin.actor AND isin.mid IN (SELECT
mid FROM isin WHERE actor = 245)ORDER BY actors.aid;
```

```
+-----+--------------+-------------+
| aid | firstname    | lastname    |
+-----+--------------+-------------+
|  13 | Johnny       | Depp        |
|  22 | John         | Candy       |
|  26 | Dan          | Aykroyd     |
|  51 | Steve        | Carrell     |
|  67 | Chevy        | Chase       |
|  76 | Sigourney    | Weaver      |
| 102 | David Hyde   | Pierce      |
| 122 | Steve        | Martin      |
| 182 | Alan         | Arkin       |
| 245 | Bill         | Murray      |
| 264 | James        | Caan        |
| 304 | Jeffrey      | Jones       |
| 307 | Sarah Jessica | Parker     |
| 308 | Martin       | Landau      |
| 309 | Patricia     | Arquette    |
| 315 | Jessica      | Lange       |
| 342 | Rick         | Moranis     |
| 351 | Anjelica     | Huston      |
| 400 | Sharon       | Stone       |
| 480 | James        | Belushi     |
| 555 | Laurence     | Fishburne   |
| 569 | Harold       | Ramis       |
| 570 | Annie        | Potts       |
| 584 | Ron          | Howard      |
| 585 | Chris        | Rock        |
| 635 | Rodney       | Dangerfield |
| 636 | Ted          | Knight      |
| 646 | John         | Larroquette |
| 647 | Judge        | Reinhold    |
| 648 | Dwayne       | Johnson     |
| 649 | Owen         | Wilson      |
+-----+--------------+-------------+
31 rows in set (1.78 sec)
```

To find the list of actors who acted with both of them, I used the aid's.

```
mysql> SELECT * FROM actors WHERE aid IN
(9,22,111,122,128,139,142,143,144,145,194,198,213,252,555,650) AND aid
IN
(13,22,26,51,67,76,102,122,182,245,264,304,307,308,309,315,342,351,400,
480,555,569,570,584,585,635,636,646,647,648,649);
+-----+-----------+-----------+
| aid | firstname | lastname  |
+-----+-----------+-----------+
|  22 | John      | Candy     |
| 122 | Steve     | Martin    |
| 555 | Laurence  | Fishburne |
+-----+-----------+-----------+
3 rows in set (0.00 sec)
```

Therefore Kevin Bacon has a two hand shake path with Bill Murray in three different cases, because actors John Candy, Steve Martin and Laurence Fishburne had shared screen space with both Kevin Bacon and Bill Murray.

5. Write a paragraph about your Semester Project including where the data is coming from.

Human genes BRCA1 and BRCA2 is responsible for the production of tumor suppressor proteins, but if a woman inherits a particular type of mutation in any one of these genes, then the person is at a higher risk of getting breast and/or ovarian cancer. Therefore, to further understand the type of mutation, I would like to create a cancer risk database which gets a nucleotide sequence as input from the user and tells us if the sequence is cancerous, its type of mutation if cancerous, its location, the change in protein, its classification.

Here is an article that is based on TP53 gene.

The IARC TP53 database: new online mutation analysis and recommendations to users. Olivier M[1], Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P.

Here is the link that has the database for both BRCA1 and BRCA2 genes.

http://arup.utah.edu/database/BRCA/Home/BRCA1_Landing