Lab 2

Daniel Ocampo

September 23, 2016

Questions

1. What is the relation schema for each data set that describes the association between their entities?

2. How are the column headers helpful for designing your schemas?

   The column help decide what tables I would need to make, kind of like my database architecture.

3. What is an appropriate primary key for each data set?

   The primary key should be entry because there i Think the data is one of a kind.

4. What kinds of memory (VARCHAR memory allocation) do you think you need to create your base? Why?

   I think we need a not null because we need an input for all of these data because we are reading it through a file.

Part 2

1. Write a query that will return the count of elements in the Entry columns in both tables.

   SELECT(SELECT count(entry) from Apop) + (SELECT count(entry) from Park);

2. Write a query that will return the distinct count elements in the Entry column. select count(distinct(Entry)) from Apop;

3. Discuss: From the above two queries, is this column a good primary key for each table? why or why not? (if not, then what column would you recommend, instead?)

   Yes because everything seems to be unique if it wasnt we would have some problems finding a specific item in the database.

4. Write a query that will return the number of records associated with Zea mays(Maize) in both tables. You might want to first determine what where entity is found in the table to create your query.

   - The entity is found in the orginism section and the command that is needed for it is SELECT count(distinct(Entry)) From Park where Organism == "Zea mays (Maize)"; 9
   - SELECT count(distinct(Entry)) From Apop where Organism == "Zea mays (Maize)"; 13

5. Write a query that will report how many organisms were listed in each table.

   - SELECT(SELECT count(Organism) from Apop) + (SELECT count(organism) from Park); 227450
   - SELECT count(Organism) from Apop ; 100082
   - SELECT count(Organism) from Park ; 127368

6. Write a query that will return the number of organisms which are common to both tables (as in, the intersection of the tables for this attribute).

   - SELECT count (*) FROM (SELECT Organism from Apop INTERSECT SELECT Organism FROM Park);
     147

7. Write a query that will return the number of proteins which are common to Apoptosis and Parkinsons, which are associated to the Zea mays (Maize) organism.

   So what was my orginal Idea was to find what both of them had common with zea mays because for some reason I thought the protein had zea mays. When in fact the organsinms had zea mays. It must of been my lack of biology.

   SELECT count (*) FROM (SELECT Proteinname from Apop where Proteinname== "Zea mays" INTERSECT SELECT Organism FROM Park where Proteinname=="Zea mays");
   4
   After getting help I discoverd that we can use random variables that used to distinguse tables in a way. For example in the bottom SQL command with a variable it can be any variable at that. I feel like we declare it after wich is why I was A little confused about it. So we are kind of like setting Apop to a and Park to p. Now we can use it to distinguish between the two tables. For comparison reasons we use the random variable because it gives us the ability to compare.

   select count(distinct(a.Proteinname)) from Apop a, Park p
   where a.Proteinname == p.Proteinname and a.organism = "Zea mays (Maize)";

8. Modify this query to print the first 15 proteins (Entry entites), according to Zea mays (Maize) are related.
   <sqlite> Select a.Entry, p.Entry from Apop a,Park p where a.Organism ="Zea mays (Maize)" LIMIT 15; A0A096QZ88—A0A061IEP5
   A0A096TM17—A0A061IEP5
   B6STV7—A0A061IEP5
   B6SUW1—A0A061IEP5
   B6T3A3—A0A061IEP5
   B6T3H2—A0A061IEP5
   B6T9F6—A0A061IEP5
   B6TB21—A0A061IEP5
   B6TZ68—A0A061IEP5
   C4J813—A0A061IEP5
   O81214—A0A061IEP5
   Q6PLR8—A0A061IEP5
   Q6PLR9—A0A061IEP5
   A0A096QZ88—A0A067RA11
   A0A096TM17—A0A067RA11
   <sqlite> Select a.Entry, p.Entry from Apop a,Park p where p.Organism ="Zea mays (Maize)" LIMIT 15; A0A009GCL8—C0JR65
   A0A009GCL8—C0JR66
   A0A009GCL8—C0JR67
   A0A009GCL8—C0JR68
   A0A009GCL8—C0JR69
   A0A009GCL8—I6WV73
   A0A009GCL8—I6XGM5
   A0A009GCL8—K7S1X3
   A0A009GCL8—Q9T6S0
   A0A009GLJ9—C0JR65
   A0A009GLJ9—C0JR66
   A0A009GLJ9—C0JR67
   A0A009GLJ9—C0JR68
   A0A009GLJ9—C0JR69
   A0A009GLJ9—I6WV73

9. Create a query to determine how many genes are in common in both tables, for all organisms

(i.e., the intersection of all information about genes across all organisms).

select count( ) from Apop a, Park p where a.gene_names ==P.gene_names and a.Organism == p.Organism;

11673

10. Create another query to determine the names of first ten of these genes which are at the intersection of both tables, across all organisms.

    select a.gene_names,p.gene_names,a.Organism,p.Organism from Park p , Apop a where p.gene_names = a.gene_names and a.Organism = p.Organism limit 15 ;

    PARK2 PRKN—PARK2 PRKN—Homo sapiens (Human)—Homo sapiens (Human)
    LRRK2 PARK8—LRRK2 PARK8—Homo sapiens (Human)—Homo sapiens (Human)
    MAPT MAPTL MTBT1 TAU—MAPT MAPTL MTBT1 TAU—Homo sapiens (Human)—Homo sapiens (Human)
    SQSTM1 ORCA OSIL—SQSTM1 ORCA OSIL—Homo sapiens (Human)—Homo sapiens (Human)
    HTRA2 OMI PRSS25—HTRA2 OMI PRSS25—Homo sapiens (Human)—Homo sapiens (Human)
    AIMP2 JTV1 PRO0992—AIMP2 JTV1 PRO0992—Homo sapiens (Human)—Homo sapiens (Human)
    Btk Bpk—Btk Bpk—Mus musculus (Mouse)—Mus musculus (Mouse)
    CDK5 CDKN5—CDK5 CDKN5—Homo sapiens (Human)—Homo sapiens (Human)
    Ddit4 Dig2 Redd1 Rtp801—Ddit4 Dig2 Redd1 Rtp801—Mus musculus (Mouse)—Mus musculus (Mouse)
    Ddit4 Redd1 Rtp801—Ddit4 Redd1 Rtp801—Rattus norvegicus (Rat)—Rattus norvegicus (Rat)
    DDIT4 REDD1 RTP801—DDIT4 REDD1 RTP801—Homo sapiens (Human)—Homo sapiens (Human)
    Bm1_48140—Bm1_48140—Brugia malayi (Filarial nematode worm)—Brugia malayi (Filarial nematode worm)
    Rattus norvegicus (Rat)—Rattus norvegicus (Rat)
    Rattus norvegicus (Rat)—Rattus norvegicus (Rat)
    Rattus norvegicus (Rat)—Rattus norvegicus (Rat)