

Data Storage and Retrieval – Assignment 5

Aaron Niskin

September 27, 2016

1. Seven Databases, Section 2.4, Day 3, Do problems 1 and 2. For problem 2, get two or more user comments from each student in the class.

1. Create a stored procedure where you can input a movie title or actor's name you like, and it will return the top five suggestions based on either movies the actor has starred in or films with similar genres.

```
CREATE OR REPLACE FUNCTION search_by_name_or_title( searchStr text )  
RETURNS TEXT AS $$
```

2. Expand the movies database to track user comments and extract keywords (minus English stopwords). Cross-reference these keywords with actors' last names, and try to find the most talked about actors.
2. Discuss whether hubway.trips is normalized.
3. Finish my solution to problem 2d in Assignment 2 (which I wrote in the assignment doc).
 - a. Using the temporary tables in the solution, write sql to make a table with columns charge, homeless arrest rate, non-homeless arrest rate, difference. Arrest rate is number of arrests for the charge divided by total arrests. Difference is the difference between rates, homeless vs. not.
 - b. For the charge with the biggest difference in favor of homeless arrest rate, show the homeless rate and non-homeless arrest rates for each year.
4. Use python with psychopg2 and the matplotlib library to create a histogram of movies by number of words in the title.