Nadim Sarras

April 27, 2017

Online Publication Management System Report

Designing a database for over a million lines of data regarding publications in Computer Engineering proved to be no easy task. However, after taking into account the amount of data, the different variables, and efficiency needed to look up data a sufficient design was established. According to the input data file, many input parameters had to be accounted for when constructing the database. These variables included; journal, pages, id, year, title, and the authors of every publication. However, since there was a possibility of multiple authors per publication, it was impossible to upload all the details of the publication into a single row. Therefore, two tables were constructed. One table called “pubs” would include most of the basic information regarding a publication including the entities; journal, pages, id, year, and title. The other table “authors” includes every author of a publication in a separate tuple followed by another entity “auth_id” which is the foreign key for ID in the “pubs” table. With this design multiple authors can be placed in multiple rows, while still preserving the specific publication that each author has contributed to. Since the publication id is used as a foreign key, any data regarding the publication can be retrieved from either of the two tables. Figure 1 below displays the design of the database. There is a many to many relationship between the pubs and authors tables since many publications can have many authors or vice versa.

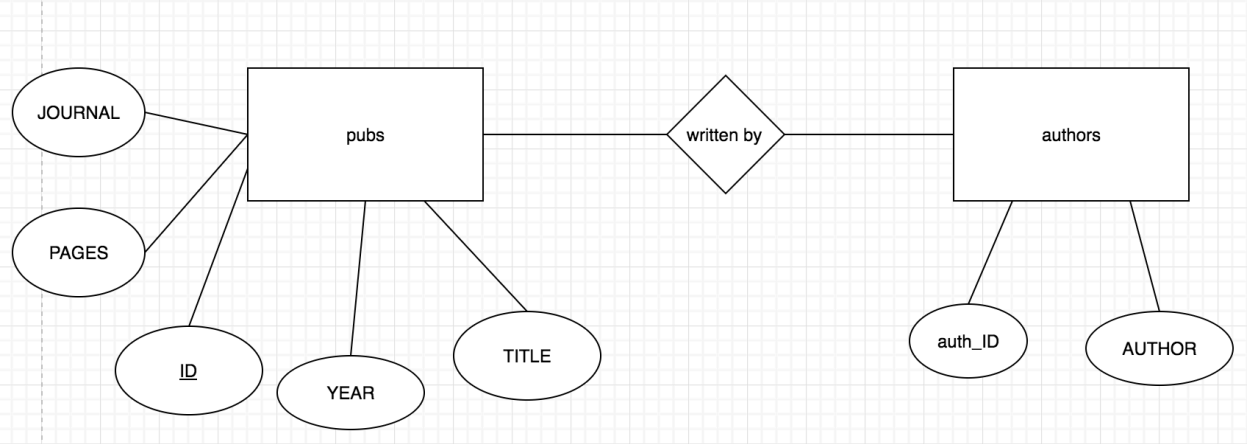

Figure 1: Publication Database ER Diagram

Furthermore, with respect to the implantation and testing of the database, various queries were implemented as well as service requests. A wide ranging multitude of queries were executed to test the database by adding and deleting tuples, testing joins, and returning data based on a variety of conditions, as specified in the notebook.ipynb file. Service requests were also used to test the functionality of the database, where the user would specify the request command and receive the return of the command in json format.

In conclusion, designing, implementing, and testing a database to manage online publications has proven to be extremely useful within the information industry and requires a broad range of technical knowledge.