

Python Page Ranker

This is a set of programs that emulate some of the functions of a search engine. They store their data in a SQLITE3 database named 'data.sqlite'.

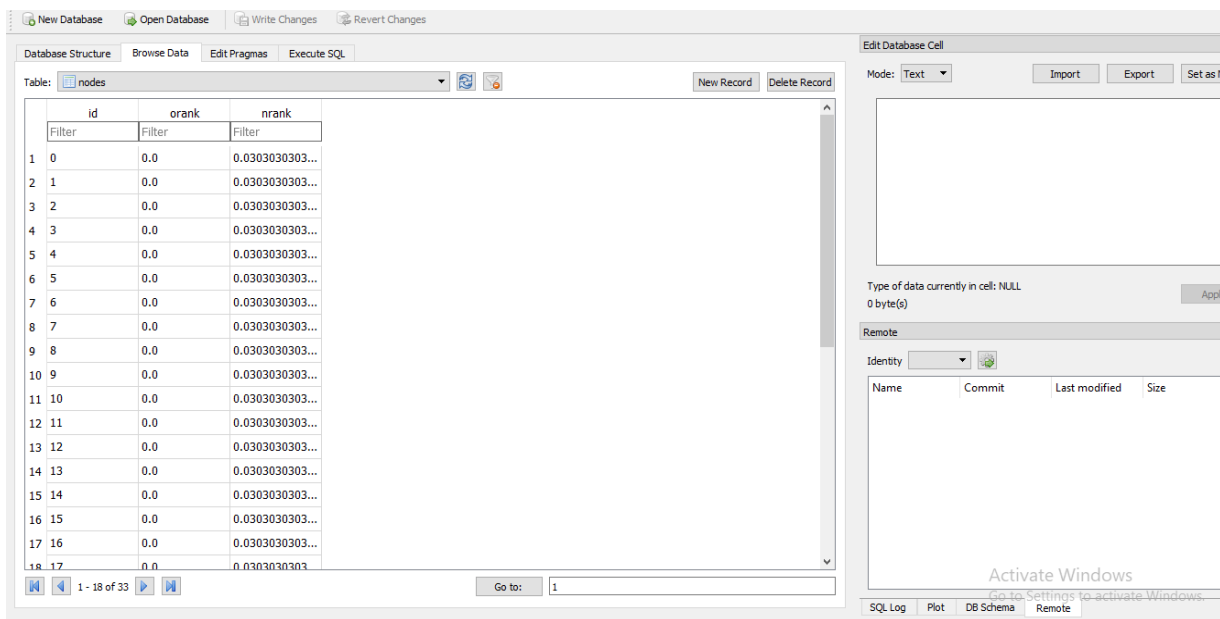
You should install the SQLite browser to view and modify the databases from.

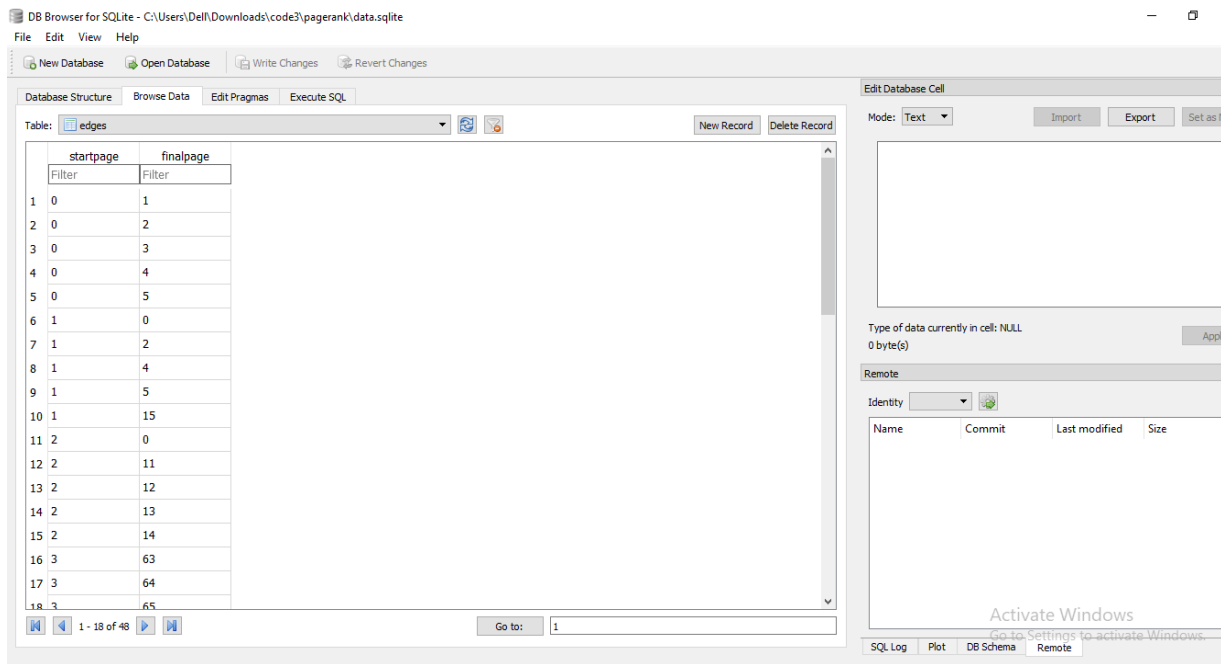
1st command to run:-

```
python make_data.py
```

this will

1. create table named nodes with attributes id(for storing page id), orank(for storing previous rank of page), and nrank(for storing current rank of page)
2. create table named edges with attributes startpage(for storing id for page from where edge starts), finalpage(for storing id for page that edge is pointing to)
3. create file data.sqlite containing all these table information





Once you have a data in the database, you can run Page Rank on the data using command:- `python calculate_rankmatrix.py`

then it will ask us to input `r=` number of iterations we want to calculate rank
How many iterations:??

```
C:\Users\Dell\Downloads\code3\pagerank>python make_data.py
C:\Users\Dell\Downloads\code3\pagerank>python calculate_rankmatrix.py
Iterations??5
138 0.003633024459470749
138 0.0011746994951603005
138 0.0003537593903298709
138 7.967997718218041e-05
138 1.9588010316739666e-05
138 [(0, 0.031111417500133666), (1, 0.02681767794787845), (2, 0.031111417500133666), (3, 0.02681767794787845), (4, 0.031111417500133666), (5, 0.05161317849556846), (6, 0.0088566171435106), (7, 0.04791165023869901), (8, 0.04256489838836883), (10, 0.030888566171435106), (15, 0.026129019478471467), (11, 0.02681767794787845), (12, 0.027794787845), (13, 0.03362528153610099), (14, 0.02681767794787845), (63, 0.026129019478471467), (64, 0.026129019478471467), (65, 0.026129019478471467), (66, 0.0261278471467), (67, 0.026129019478471467), (16, 0.02681767794787845), (17, 0.02681767794787845), (18, 0.02681767794787845), (19, 0.02681767794787845), (9, 0.042564898383), (20, 0.02665853719331689), (30, 0.029496354876049326), (290, 0.029496354876049326), (21, 0.028642883514438784), (22, 0.028642883514438784), (23, 0.028642883514438784), (24, 0.028642883514438784), (137, 0.02986070870471732)]
C:\Users\Dell\Downloads\code3\pagerank>
```

For each iteration of the page rank algorithm it prints the average change per page of the page rank. The network initially is quite unbalanced and so the individual page ranks are changing wildly. But in a few short iterations, the page rank converges. You should input iterations large enough such that the page ranks converge.

You can run `calculate_rankmatrix.py` as many times as you like and it will simply refine the page rank the more times you run it.

If you want to restart the Page Rank calculations, run command :-
`python reset_rankmatrix.py`

It re-initializes all pages and set to a rank of $1.0/N$. Then prints message "All pages set to a rank of 1.0" on completion

If you want to visualize the current top pages in terms of page rank,
run command :-
`python make_graph.py`