Coursework Group 1

Group Members
Olivia Wilson
Tianhua Mu
Zhen Wang
Weijian Yao
Yuhe Chen

FoDS(COMP6235)

Github Repo: https://github.com/academic-jobs/academic-jobs

Justification

- An application which helps people find trends in academic jobs data.

- Sources: jobs hosted on www.jobs.ac.uk and REF results from 2014

- Potential impact of REF on job hiring

- Could provide an insight into the inner workings of university hiring

Data collection - Job scraping

- scrapy Python module
- This was used because of the broad built-in functionality
- Sub-pipeline within scrapy:
  - 'Spiders' all relevant links from www.jobs.ac.uk
  - If already downloaded then skip
  - Otherwise, download page and scrape specified details
  - **Clean data**
    - Parse datetime
    - Clean-up whitespace
    - Field names etc
  - Insert into MySQL database

Data Collection - REF

- Research Excellence Framework
- Controls funding provided to universities - basically proxy for 'prosperity'/'quality' of a department
- Provided by HEFCE at www.ref.ac.uk

- Simple CSV file
- 'Quality scores' are provided per *university department*
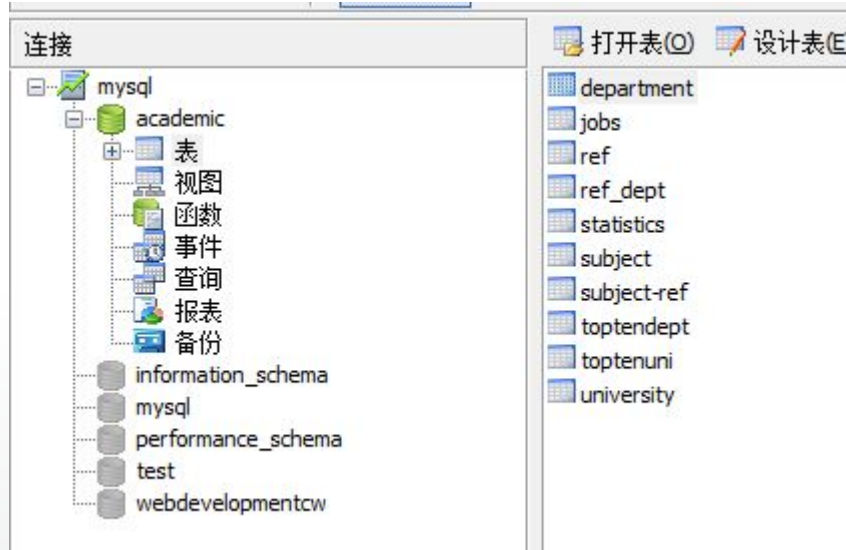
MySQL Database

Demo!

Back End

- – jFinal framework
- – MVC pattern
- – JDBC connect to MySQL database

Demo!

Analysis Function
//count numbers of job of each university
 private List<Jobs> countjobs(int a)
//count job_types for each university
   private List<Jobs> counttype(int a)
// top10 departments
   private void toptendepartment()
// top10 university
   private void toptenuniversity()
//sum all university stars for each department
   private float [] sumdeptstar(int l)

Demo!

Front End

Several web pages to display our consequences:

Characteristics:

– Interface:HTML5 and CSS3

– Logic: Javascript

– Interactive

– Responsive

Responsive

Use bootstrap

Change the size according to the size of window

Suitable for more device:

– mobile

– ipad

UNIVERSITY OF
Southampton

Demo Shiny Server

UNIVERSITY OF
Southampton

Conclusions

- There is a small positive correlation between job counts and REF score
- There are very little universities that have low REF scores with higher job counts

Further Work

- Continue scraping to get a better time series of data
- Clean data so that university name is invariant to "the"

Thanks!