Day 22: Finishing up PiCloud work and preview of Juriscraper workshop

Raymond Yee

April 11, 2013 (`http://bit.ly/wwod1322`)

Agenda / Homework

Continue on the problem set due on Monday. Fill in
`segment_status` and run the notebook, showing the number of
cores vs time for your run on PiCloud. Send me a URL to a public
notebook (rendered in nbviewer.ipython.org). Due Monday, April
15, 2013 at noon. e.g., my notebook in my PiCloud account

Follow up problem:

- ▶ how to calculate the cost of the jobs on PiCloud
- ▶ how to extract the pdfs in the corpus? How much will it cost?

Juriscraper Workshop

Please let me know whether you will attend our Workshop on
JuriScraper on Friday, April 12, 2013 in Room 205 South Hall (1-3
pm with 3-4pm for die-hards.) This is a great opportunity to learn
about web scraping. Mike Lissner will be there along with Brian
Carver.

To retrieve a notebook we will use tomorrow:

```
import urllib
urllib.urlretrieve ("http://bit.ly/jsnotebook", "Day_22_ju
```

Next Steps

- Focus will be on projects.
- I'd like us to work on the problems that come up while you make progress on projects. For example, plotting counties, brought up because of student project.
- I'm also interested in helping you work with data in the longer term...that's why I think projects are important.
- Final products will be iPython notebooks that you write to be integrated into the Working Open Data repository – we'll use `git pull requests` and a community review process similar to what Fernando discussed. (See https://vimeo.com/63250251 for a video of his PyData 2013 keynote.)
- Final presentations will be on Tuesday, May 7 from 2-3:30pm. Invite your friends!!
- **More details next week**