Day 19: PiCloud, CommonCrawl revisited, Projects

Raymond Yee

April 2, 2013 (`http://bit.ly/wwod1319`)

Agenda

- Review of CommonCrawl
- Fernando Perez will be speaking to us on Thursday.

Midterm

- See midterm key
- Scaling of grades: anything above score $+$ (score-60)/4 $->$ if you had 84, you would get 66/60

CommonCrawl Review

- On Day 18, Dave Lester gave a nice talk about Common Crawl: http://www.slideshare.net/davelester/introduction-to-common-crawl

- Today, we will work through: Day_19_CC_etc.ipynb

Homework

- Let me know whether you are close to running out of credit on http://picloud.com. We want to be ready to work on CommonCrawl.org using PiCloud.com.
- I will develop a homework set to lead you through calculating stats on the CommonCrawl data set.