**In person meeting summary**

Attended:

Anthony, Alex L, Jiaro, Joe, Tony

Missed:

Alex B

Work

**Ideas from Meeting**

Web client architecture

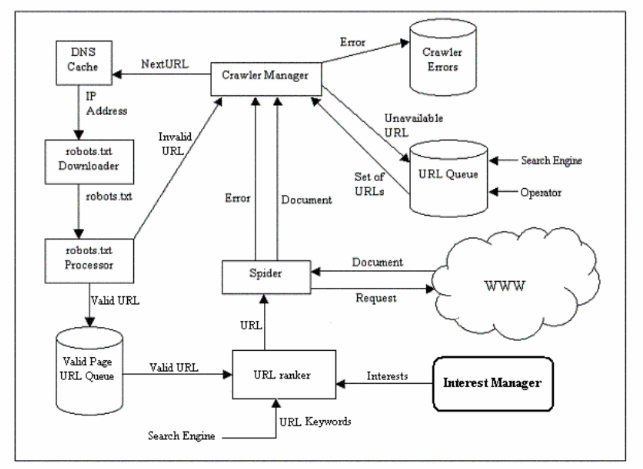
Notes

* Politeness Policy
  + <http://www.robotstxt.org/robotstxt.html>
* Politeness Policy in an architecture
  + <https://www.researchgate.net/figure/Architecture-of-a-Centralized-Regional-Crawler-Page-ranking-crawlers-crawl-the-pages_fig3_255647292>
* Architecture of a web crawler
  + <https://www.researchgate.net/publication/313443438_Architecture_of_a_WebCrawler>
  + <http://www.cs.sfu.ca/CourseCentral/456/jpei/web%20slides/L05%20-%20Web%20crawling.pdf>
  + Useful diagrams in this resource: <http://shodhganga.inflibnet.ac.in/bitstream/10603/2415/13/13_chapter%204.pdf>

Session Table as stored in Database

* Session number
* Date
* Time
* URL List

Web crawler scraping

* Show content of certain search
  + Grab link
  + Manually place html element onto scraper
    - Everything of page is scrapped
      * Dictionary
      * Songs
      * Recursive back into scrapper
* Returns list of pages
* We would extend the scaping to iterate the list of pages

Octoparse

* https://www.octoparse.com/blog/top-20-web-crawling-tools-for-extracting-web-data

Scappy