Minutes:

* Participants: Chun, Max, Ali, Zhen and Alan.
* \* Meeting next week 1-4 pm.
* \* First Meeting place will be bv473
  + look for room in library or outside
  + Third meeting at the place behind AC220 from 1:20 to 2:30pm.
* \* No one knows about git, it should be easy to learn
* \* By next week, in meeting we can talk about it, and see how everyone is doing. Everyone is good about the git.
* \* No one understands the assignment before the first week. Now we have a clear image of the assignment.
  + Researcher: Look for a key word from an article and search for that key word in the target websites. Then, store all the articles that contain the key word in the database, including all the aspects which have been talked in schema.

Runs one time per day - Ali

* \* Three main components of the assignment
  + Parser, database and UI, test cases
    - Difficulty levels
      * Parser 5-7
    - Cherry py, pygame
    - Ask professor if mongo db is good or is there another one for this project
    - UI doesn’t matter which language it is on
  + Comfortable with python
* \* If not database, we can use a text file for storing data. We finally decide to use Mongo DB to store data.
* \* Platforms: mostly comfortable with python
* \* Talk about regex in next week meeting
* \* We have worked with regex before, parsing through html code in C
* \* Database, talk about different database and the benefits each one
* \* Worked with java with something about hospital, Zhen
* \* Login and search the patients, Max/Zhen
* \* Easy to make UI in VB, Chun
* \* Have not worked with javascript before,
* \* Still need to decide on UI, Ali
* \* Ask questions in tutorial,
  + What questions do we wanna ask,
  + Users
    - We gotta talk about the two components of the assignment, twitter and news articles
    - What kinda data does the librarian want? Max
    - What kinda UI does the anthropologist want,
    - what kinda input? And what sort of output is he expecting?
    - How much data should we display / graph he wants to look at? Chun
    - What options does he want in the program
    - Do you have an example of what this program should do
    - Does the program look at the web pages everyday or every hour
  + Professor
    - Does the professor want to parse only news article or the twitter accounts as well (twitter vs news website)
* \* Suggestion made by Chun, comment every line,
* \* Work out the standard on coding in the next meeting?
* \* More question for the anthropologist, should we working on both, like twitter or website? Any specifics, choose one?
* \* Should the program be running constantly? Run it on a schedule? Is it going to be on a server?
* \* Talk about the schema next week?
* Translations?
  + News website
    - Name
    - Homepage
    - Region/area of coverage
    - article
      * Link
      * html
      * External links
      * Date
      * Author
      * Article title
      * Text
      * Quotes
        + Name
        + Text
        + Link
      * Hits (internal)
* \* Draw a image for schema, like student number etc? Chung
* \* Have a solid schema before actually coding or implementing?
* \* Ideas:
* o Cool interface, very user friendly
* o User manual, Cheng
* o Searches, like advanced search, filters etc
* \* Python webkitwebview, use for parsing,
* \* Beautfiulsoup, for getting anchor tags etc,

Database – Schema is good, do it on Mongo DB

* Parser – 3 components

\*Extracting html source

\*Key word in html

\*Check the entry in the database or not

Beautiful soup – Python

Database runs locally on the computer, so make one folder with database in dropbox and run it

Separate the work

UI – Zhen and Chun, cherry py

Parser – Max, Alan

Database – Ali

What we going to do before next meeting:

Install Mongo DB; learn cherry py and Beautiful Soup; google Mongo engine

Find software which can track user stories