Iteration Cycle 2

Date: 10.31-11.02

Team member: Bangjie Liu, Weixiang Ding, Qing Li, Yi Wang, Wenyan Hu.

* Discussed with team members to specify the target website.

The target website for crawler should maintain following features:

1. No restrictions to get link or addresses
2. No restrictions for downloading images
3. No restrictions for connections from the same IP address
4. No restrictions to parse the html content.

Alternate Target website:

1. New York Times (http://www.nytimes.com/)
2. Imdb (<http://www.imdb.com/>)

* Unify and configure the project development environment

Coding: python 3.4

Database: MySQL

* Configure the GitLab for every team member

[http://stackoverflow.com/questions/3860112/multiple-github-accounts-on-the-same-computer](http://stackoverflow.com/questions/3860112/multiple-github-accounts-on-the-same-computer" \t "_blank)

Classify and manager the GitLab project file for code and docs.

* Tasks for Cycle3:

Web Crawler mini team : Yi Wang, Weixiang Ding, Wenyan Hu

1. Get familiar with python, crawler framework & tools.
2. Analysis target website HTML structure and source code to come up with a feasible crawler approach.

Database mini team : Bangjie Liu, Qing Li

1. Get familiar with AWS
2. Set up database on AWS and get it run.