<CS410 : Project Proposal>

* Team name : guava
* Team members :

1. Hongseok ha (netID : hh23), with administrative duties
2. Changsoo Kim (netID : ck37)

* Project topic : (Option 1) Reproducing a paper: Latent Aspect Rating Analysis
* Programming language that will be used : Python
* Dataset Availability
  + Two datasets (the hotel review dataset and the Amazon Mp3 review dataset) that are used in the paper are both available in the link below.
    - Dataset Link : <http://sifaka.cs.uiuc.edu/~wang296/Data/index.html>
  + However, for the Amazon MP3 review dataset, the number of reviews are different in the paper and the actual dataset. (the paper : #16,680, the dataset : #55,740) We will see if we can get results similar to the paper.

The paper (Latent Aspect Rating Analysis) deals with the methodology of mining latent topical aspects, without pre-specification of aspect keywords (hashtags, … ). LARAM would be helpful to cluster user reviews with various aspects, making it easy for other users to retrieve topics that they want. Also, reproducing this paper would be helpful for us to understand overall contents of CS410.