CS410 Text Information Systems
Project Proposal and Team Formation

We are the team *American River*. The team members are:

- Maximus Choi (daechoi2@illinois.edu)
- Jackson DeDobbelaere (dedobbe2@illinois.edu) – Captain
- Subramanian Shankar (ss117@illinois.edu)

For the project, we will reproduce the paper "Latent aspect rating analysis without aspect keyword supervision" by Hongning Wang, Yue Lu, and ChengXiang Zhai.

Background:
Latent Aspect Rating Analysis (LARA) is the task of inferring both opinion ratings on topical aspects, and the relative weights reviewers have placed on each aspect based on review content and the associated overall ratings. This paper proposes a model which effectively performs the Latent Aspect Rating Analysis task without the supervision of aspect keywords and explores all kinds of opinionated text data containing overall sentiment judgments and support a wide range of interesting application tasks.

Main Goal/Language to be used:
We will use the programming language of *Python* to reproduce the model proposed in the paper.

Data:
The data used in the papers are two review datasets:

1. a data set of 235,793 hotel reviews acquired by crawling TripAdvisor in one-month period (from February 14, 2009 to March 15, 2009, and
2. a data set having information of an MP3 players crawled from www.amazon.com.

Dataset 1 acts as the ground-truth for quantitative evaluation of both aspect identification and latent aspect rating. Here, Reviewers are also asked to provide ratings on 7 pre-defined aspects in each review: value, room, location, cleanliness, check in/front desk, service, business service ranging from 1 star to 5 stars. Dataset 2 has only one overall rating in each review, ranging from 1 star to 5 stars.

Both the datasets are available in http://sifaka.cs.uiuc.edu/~wang296/Data/index.html