Week 1 Report

Michael Meding

February 23, 2015

Over the past week I have gotten in contact with Hoanh Nguyen who was my partner for the final project in NLP last semester. We discussed several of our ideas for the project for this semester. The main topic of which was to look deeper into the SemEval competition for ideas. This would be extremely beneficial because the data sets that we would need for evaluation exist freely which was a major problem that we encountered with our project last semester.

Our project from last semester was a binary classifier that identified bullies using only data from twitter. This data was both collected and annotated by Hoanh and I which was an awful but arguably necessary experience. Additionally, the results from our experiment were far less spectacular than anticipated. Which leads me to this semester. This semester Hoanh and I had discussed improving our project from last semester with both better data and far more suitable set of features. These features would likely be taken from prior SemEval competitions such as using Wille Boag's Twitter sentiment analyser. We would also be improving on the core binary aspect of our project. Bullying, in a broad sense is not a binary action. This means that there are more than just two Twitter users that are likely to be involved in a case of bullying. This improvement would likely come by restructuring of our classifier to incorporate many more aspects of bullying such as antagonists or bystanders. It would also be great to do some data visualization features for this but as the semester has been seriously foreshortened so this may not be a possibility.