Khyati Kansagra
Roshni Ramakrishnan
Aniruddh Garge

Project Progress Report

We plan on working on prediction of location from twitter data. The approach that we have decided involves five steps-

1. First, we will obtain the data of every city with every landmark that is listed on its Wikipedia page and create a sort of database for the same.
2. We then plan on building a knowledge base for the links that go from the Wikipedia page to any other location or landmark.
3. We will then generate a graph for the cities and the associated landmarks with nodes connecting to the city if a link exists.
4. We perform the processing of the tweets in this step and find an overlapping measure so that if any word is present in the graph generated in above steps the probability is high that the user has tweeted from the same city.
5. This overlapping score would be fed as an input to a classifier like SVM which would output a probability to determine the city the user has tweeted from.

This is a tentative approach and is subject to change if suggested otherwise by the TA or the professor himself.