

CSE 6339 Course Tutorial Proposal

By - Rohit Bhoopalam (Student ID: 1001100534)

I want to show some usage of Python NLTK library through the following 3 tasks,

Task 1:

- Introduction to NLTK
- Downloading and using the NLTK corpus
- Tokenizing the text
- POS tagging - Unigram tagging, Bigram tagging
- Using Named Entity Recognizer

Task 2:

Will show how to do sentimental analysis, using Movie reviews data set of nltk & nltk's Naive Bayes classifier

Difference in accuracy when training data is represented as an individual word VS most influential words + bigrams

ref: <http://streamhacker.com/2010/05/10/text-classification-sentiment-analysis-naive-bayes-classifier/>

Task 3:

The amazon tutorial for doing sentimental analysis using Mapper & Reducer with Python NLTK,

This has the following steps,

- [Step 1: Create a Twitter Developer Account](#)
- [Step 2: Create an Amazon S3 Bucket](#)
- [Step 3: Collect and Store the Sentiment Data](#)
- [Step 4: Customize the Mapper](#)
- [Step 5: Create an Amazon EMR Cluster](#)
- [Step 6: Examine the Sentiment Analysis Output](#)
- [Step 7: Clean Up](#)

This Amazon tutorial consists of pulling data from twitter, using nltk's naive bayes classifier, using mapper & reducer to process the data.