

CSCI 544: Applied Natural Language Processing

Course Project Proposals

Jonathan May and Nanyun Peng

1. What is the name of your task?
2. Why this is an interesting data set? What would be learned from building a good classifier of this data?
3. How will the data be collected and labeled?

4. How many annotated examples do you think you will be able to collect? Provide a justification for this number (e.g. note how many instances of unlabeled data are in your source repository and guess how many you will have after cleaning and filtering. Then, if your team members will do the annotation, calculate how many items they will need to annotate each week in order to hit your target).

5. What are the names and use email addresses of your team members?