

**Assignment #4.****Due December 6<sup>th</sup>, 11:59pm-midnight****Dialog analysis**

In this assignment we are going to analyze dialog text. Modern dialog text often occurs online in forums or chats, or on transcripts of phone conversations; but for this exercise we are going to analyze a theater play, because the sentences are typically longer and richer, and the text is clearly annotated with the speakers and settings.

To try to accommodate the desire for a free project as well as a more structured homework, this assignment is going to be very open ended. I will not specify what to turn in, but I will offer guidelines.

**Assignment: Pick your favorite theater play for which you can get the text and analyze it using methodologies of your choice as seen during the course. Turn in a report. Your report should discuss your methods and your findings and should contain visually appealing representations of your results.**

Grading. You will be graded solely on your report. I will consider your choice of methods: What do you do and how do you do it. Do you do it correctly? How rich is your data and how interesting are the conclusions and thoughts you derive from it, as well the appeal of the visualizations you choose to present. I will take into account your effort as well as the overall quality of the write up. A large number of facts without any explanations or conclusions does not constitute a very good report. A lot of speculation with no data to back it up is not very good either. If you work in a group it is acceptable to submit a single report for the whole group.

These are some ideas for things that you might want to analyze to get you started in your thinking. Feel free to extend, or deviate as you please. What makes most sense will depend on your choice of play.

- What are the temporal references? If your play contains actual dates (like in a diary) then relative times (like “last Thursday”) can be converted to absolute references (the Stanford NLP can do this. See this online demo: <http://nlp.stanford.edu:8080/sutime/process>).
- Topics. What are the topics of the play? You could make a break down by character. Do these topics evolve over time? (E.g. by act).
- What are the Named Entities (People, Places, Companies) that appear in your play?
- Who talks (spends time) with whom? Does it change over time? Which fraction of the talk does each speaker contribute?
- Who talks about what or whom? What is the sentiment of the speaker about this entity? (Or at least, what is the sentiment of the words that appear near them in the dialog?).  
Notice that you will probably want to do coreference resolution (to identify what pronouns matches what noun) when identifying who talks about what or whom. In dialogs, participants agree about the antecedents of pronouns, so you may want to process consecutive utterances from the various dialog participants as a single unit of text for the purpose of coreference resolution.
- Sentiment. What is the mood of each speaker? (E.g. the average sentiment of the words they utter?) Does it change over time? Does it depend on who they are talking to or whom or what they are talking about?

**Have fun!**