

# LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

6 October 2016

# Class outline

- ▶ Revision of what we discussed so far
- ▶ Discussion about final projects
- ▶ Planning for the tutorial session

# Topics discussed so far

1. Regular expressions
  2. Pre-processing of text: tokenizing, sentence splitting
  3. Text normalization, morphology: spelling correction, stemming, lemmatization
  4. Language models, evaluating them, building and using them
  5. POS tagging
- ... +3 assignments.

# Topics for the coming weeks

1. Text classification (most useful in my opinion)
2. Parsing sentences
3. Other topics: Discourse, Semantic analysis etc.
4. NLP Applications (quick overview)
5. NLP for CALL

# Specific Issues

- ▶ What has been difficult so far conceptually, apart from mathematical notation?

# Specific Issues

- ▶ What has been difficult so far conceptually, apart from mathematical notation?
- ▶ What has been difficult in terms of writing python code?

# Specific Issues

- ▶ What has been difficult so far conceptually, apart from mathematical notation?
- ▶ What has been difficult in terms of writing python code?
- ▶ Is there sufficient time to do the assignments?

# Specific Issues

- ▶ What has been difficult so far conceptually, apart from mathematical notation?
- ▶ What has been difficult in terms of writing python code?
- ▶ Is there sufficient time to do the assignments?
- ▶ What specific topics would you want me to revise in an extra class?



# Specific Issues

- ▶ What has been difficult so far conceptually, apart from mathematical notation?
- ▶ What has been difficult in terms of writing python code?
- ▶ Is there sufficient time to do the assignments?
- ▶ What specific topics would you want me to revise in an extra class?
- ▶ What specific programming practice do you want?

## Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?

## Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?
- ▶ Do you refer to any other sources beyond classroom slides and NLTK?

## Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?
- ▶ Do you refer to any other sources beyond classroom slides and NLTK?
- ▶ How many of you actually write some programs for their dissertation work or RA work?

# Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?
- ▶ Do you refer to any other sources beyond classroom slides and NLTK?
- ▶ How many of you actually write some programs for their dissertation work or RA work?
- ▶ How many of you still are considering writing some programs to use in your dissertation work?

# Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?
- ▶ Do you refer to any other sources beyond classroom slides and NLTK?
- ▶ How many of you actually write some programs for their dissertation work or RA work?
- ▶ How many of you still are considering writing some programs to use in your dissertation work?
- ▶ Do you think you have a picture of what kind of problems one faces in writing code for text processing?

# Some questions

- ▶ How many of you regularly follow NLTK book examples on topics discussed in the class?
- ▶ Do you refer to any other sources beyond classroom slides and NLTK?
- ▶ How many of you actually write some programs for their dissertation work or RA work?
- ▶ How many of you still are considering writing some programs to use in your dissertation work?
- ▶ Do you think you have a picture of what kind of problems one faces in writing code for text processing?

# Planning for a Tutorial Session

1. Any day after 17th October: Thursday or Friday evening, after 5pm
2. Weekends: Saturday will work, provided I am in town
3. What will happen: some amount of discussion and lot of group work, or the other way round?
4. Topics: let us decide before hand via a wishlist.



# About Final projects

- ▶ Some example ideas are uploaded on BB. You can come up with your own ideas, but get them approved by me.
- ▶ Carries 20 marks. 10 marks for written reports, and a oral defense. 10 marks for the actual implementation.
- ▶ Initial report describing your idea and plan: Due on 5th November.
- ▶ Final submission due: 9 December.

# Ideas

- ▶ Ideas on BB
- ▶ Anyone who already know what they want to do can mention it.
- ▶ Others without that clarity yet: Decide before we finish text classification. We will discuss this again in 2 weeks!

# Pen and paper exercise

Exercises 5.2–5.4 in J&M

- ▶ Using the tags from PTB and CLAWS7 tagsets, tag the following sentences manually, ignoring the punctuation. Compare your tags with your neighbors to discuss the agreement.
- ▶ Sentences: (you can also compare yourself to Stanford POS tagger, for example)
  1. It is a nice night.
  2. This crap game is over a garage in Fifty-second Street...
  3. ...Nobody ever takes the newspapers she sells...
  4. He is a tall, skinny guy with a long, sad, mean-looking kisser, and a mournful voice.
  5. ... I am sitting in Mindy's restaurant putting on the gefillte fish, which is a dish I am very fond of ...
  6. When a guy and a doll get to talking pecks back and forth at each other, why there you are indeed.

# Next Class

- ▶ Text Classification - Introduction
- ▶ Please read: Chapter 5 in Language and Computers. Get hold of atleast this chapter from someone in the class who has the book. No other book introduces text classification in such light weight language.
- ▶ Video: 10.4 in Radev's course. 25min video.
- ▶ For practical stuff: Chapter 6 in NLTK