LING83600: LANGUAGE TECHNOLOGY
FALL 2020
CUNY GRADUATE CENTER

Instructor: Prof. Kyle Gorman
Lecture: TBD
Office hours: TBD

SYNOPSIS

This course explores how computers process human language. Key technologies emphasized include word similarity, computational morphology and syntax, topic modeling, machine translation, and speech recognition and synthesis. Students are expected to be familiar with basic linguistic notions like phoneme, morpheme, (syntactic) head, constituent, etc., and to be comfortable developing in the Python program language.

OBJECTIVES

Students will become familiar with the basic models guiding the aforementioned technologies. They will also learn to critically read the academic literature in speech and language technology.

MATERIALS

The primary text will be Eisenstein 2019, but some additional will be assigned. Students are also welcome to use the Computational Linguistics Laboratory (7400.13) for practice and assignments.

ASSIGNMENTS

Assignments will take the form of a small software development projects accompanied by a write-up describing the general approach taken and any challenges encountered. Students will usually be able to verify the technical correctness of their code by running a provided unit test. Students will also be graded on the readability of their code, and the quality of the write-up. We will use GitHub Classroom for assignment turn-in.

The final assignment will be an open-ended project. Students are encouraged to conceive of projects relevant to their research interests. Students should discuss project plans with the instructor during office hours to confirm that it is both feasible and of appropriate scope.

GRADING

80% of students' grades will be derived from the assignments; the remaining 20% will be reserved for participation and attendance. Assignments must be submitted on time or will receive a 0 grade (barring a documented emergency).

ACCOMMODATIONS

The instructor will attempt to provide all reasonable accommodations to students upon request. If you believe you are covered under the Americans With Disabilities Act, please direct accommodations requests to Matthew G. Schoengood, Vice President for Student Affairs.

ATTENDANCE

Students are extended to attend all lectures and practica. The instructor reserves the right to tie grades to attendance records. The instructor and practicum leader are not responsible for reviewing materials missed to absence.

INTEGRITY

In line with the Student Handbook policies on plagiarism, students are expected to complete their own work. However, a student is permitted to collaborate with another student during the coding phase of an assignment so long as they: do not share lines of code with each other, *mutually* disclose their collaboration in their write-ups, and do not collaborate *at all* on their write-ups.

The instructor reserves the right to refer violations to the Academic Integrity Officer.

RESPECT

For the sake of the privacy, students are asked not to record lectures. Students are expected to be considerate of your peers and to treat them with respect during class discussions.

SCHEDULE

(Dates TBD.)

 Word frequency distributions
 Computational lexical semantics
 Topic modeling
 Computational phonology & morphology
 Constituency parsing
 Dependency parsing
 Neural classifiers
 Neural sequence models
 Machine translation
 Speech recognition
 Speech synthesis

BIBLIOGRAPHY

Baroni, M.. 2009. Distributions in text. In Lüdeling, A. and M. Kyöto, M. (ed.), *Corpus linguistics: An international handbook*, pages 803-821. De Gruyter.

Bird, S., Klein, E., and Loper, E. n.d. *Natural language processing with Python*. URL: https://www.nltk.org/book/.

Chan, E.. 2008. Structures and distributions in morphology learning. Doctoral dissertation, University of Pennsylvania.

Eisenstein, J. 2019. *Introduction to natural language processing*. MIT Press

Goldberg, Y. 2017. *Neural network methods for natural language processing*. Morgan & Claypool.

Gorman, K., McCarthy, A. D., Cotterell, R., Vylomova, E., Silfverberg, M., and Markowska, M. 2019. Weird inflects but OK: making sense of morphological generation errors. In *Proceedings of the 23rd Conference on Computational Natural Language Learning*, pages 140-151.

Jelinek, F. 1997. *Statistical methods for speech recognition*. MIT Press.

Jurafsky, D., and Martin, J. H. 2009. *Speech and language processing*. 2nd edition. Pearson. (See also: 3rd edition draft.)

Klein, D. and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423-430.

Kübler, S. and McDonald, R. and Nivre, J. 2009. *Dependency parsing*. Morgan & Claypool.

Yang, C.. 2013. Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language?. *Significance* 10:29-34.