

Linguistics
LING 83800: Methods in Computational Linguistics II
Spring, 2020

Instructor: Prof. Michael Mandel

Email mim@sci.brooklyn.cuny.edu

Phone 718-951-5000 x2053

Web <http://mr-pc.org>

Office Graduate Center: Room 4327, Brooklyn College: 2232 Ingersoll

Office hours by appointment

Course meetings Fridays 11:45am–1:45pm, in Graduate Center room TBD

Synopsis This course is the second of a two-semester series introducing computational linguistics and software development. The intended audience is students interested in speech and language processing technologies, though the materials will be beneficial to all language researchers.

Learning goals Using the Python programming language, students will learn core algorithms used to build speech and language technologies, and best practices for evaluation and basic statistical analysis.

Topic list Topics may include, but are not limited to:

1. Probability
2. Git
3. Formal languages
4. Finite automata
5. Language models
6. Finite-state grammars
7. Generative classifiers
8. Hidden Markov models
9. Discriminative classification
10. Evaluation
11. Descriptive data analysis
12. Inferential data analysis

Materials Chapters from several textbooks will be provided as reading assignments in addition to several published papers. Students are strongly encouraged to bring a laptop computer to the practicum.

Assessment Assessment will be based on three main components: homework assignments throughout the semester, a final project, and participation and attendance in class meetings.

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

The final project will be an open-ended project which will either extend earlier homeworks or build and evaluate a speech and language technology system. Students are encouraged to conceive of projects relevant to their research interests. Students should discuss project plans with the instructor to confirm that it is both feasible and of appropriate scope.

The final grade in the course will be curved, weighting the above components as follows:

Assignment		Assesses
Participation / attendance	10%	Theoretical and applied topics
Homework assignments (×4)	15%	Applied topics
Final project presentation	10%	Applied topics
Final project paper	20%	Applied topics

Online Resources Slides, assignments, and readings will be posted on the course website:

<http://mr-pc.org/t/ling83800/>

The course will also have a **Blackboard** site for distributing readings.

Grading All homeworks and projects should be turned in via github classroom at least 30 minutes prior to the beginning of the corresponding class period. Homeworks turned in late will be penalized 10% for each day they are late. A project that is turned in two days late and would have received a 100% will instead receive an 80%.

University policy on Academic Integrity The faculty and administration of CUNY support an environment free from cheating and plagiarism. Each student is responsible for being aware of what constitutes cheating and plagiarism and for avoiding both. The complete text of the CUNY Academic Integrity Policy can be found at <http://web.cuny.edu/academics/info-central/policies/academic-integrity.pdf>. CUNY guidelines for avoiding and detecting plagiarism can be found at http://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Policies/General/AvoidingPlagiarism.pdf. If a faculty member suspects a violation of academic integrity and, upon investigation, confirms that violation, or if the student admits the violation, the faculty member MUST report the violation.

Course policy on Academic Integrity While you are encouraged to discuss the course material and assignments with your classmates and anyone else you might like, all submitted individual assignments must be *strictly your own work*. If you incorporate any work from other sources, including existing web pages, publications, books, or conversations, it should be explicitly cited with proper credit given to the original author. In the case of copying, both the copier and the copy-ee may be equally guilty.

You may work in small groups for the final project, in which case the groups will be established early in the term and each group will turn in a single final project report and present a single project presentation.