General Information
CSC/LIN 205, “Computational Linguistics”
Department of Computer Science
Grinnell College
August 26, 2016

In CSC/LIN 205, we’ll study computational techniques for producing and processing text in natural languages and the theoretical basis for those techniques, both in linguistics and in computer science.

The class meets in Noyce 3819 at 10 a.m. on Mondays, Wednesdays, and Fridays, beginning on Friday, August 26.

The Textbook

Here are the Web links for the book itself and the authors:

• https://mitpress.mit.edu/books/foundations-statistical-natural-language-processing
• http://nlp.stanford.edu/~manning/
• http://www.cis.uni-muenchen.de/schuetze/

For our programming work, we’ll use the Scheme programming language, as specified by the Revised^7 Report on the Algorithmic Language Scheme.

• http://trac.sacrideo.us/wg/raw-attachment/wiki/WikiStart/r7rs.pdf
• http://trac.sacrideo.us/wg/wiki/R7RSSmallErrata

The Instructor
The instructor for this course is John David Stone. My office is Noyce 3829, near the east end of the long corridor on the third floor of the Noyce Science Center, on the north side (facing Eighth Avenue). My telephone extension on the Grinnell College campus is 3181.

My office hours for fall 2016 are

• Mondays, 2 to 4 p.m.
• Tuesdays, 10 a.m. to noon
• Thursdays, 2 to 4 p.m.

or by appointment.

I can be reached by telephone at extension 3181.

If you use PGP encryption, you can also send me e-mail as reseda@grinnell.edu. The PGP public key associated with this e-mail address is available on the Web at

https://unity.homelinux.net/reseda-public-key.txt

and the “fingerprint” that you can use for confirmation is

EA7D 19D0 1A17 28D8 8F88 F745 6BC0 36C6 CDFD A8C2

If you haven’t yet started encrypting your e-mail, the Free Software Foundation’s guide, Email Self-Defense (at https://emailselfdefense.fsf.org/) explains how to do so.
Requirements

Each student in the course is expected to read the assigned handouts carefully, to learn the ideas, methods, and techniques presented there, to submit solutions to exercises requiring the application of those ideas, methods, and techniques, to prepare for and attend the sessions of the class, and to take and pass an examination at the end of the course.

In the schedule of topics for the course, I have specified a reading assignment between successive class meetings. Please study each specified section before the beginning of the class that follows it in the schedule.

Exercises

From time to time, I’ll propose exercises, sometimes requiring prose solutions and sometimes requiring the development of a computer program that meets certain specifications.

A deadline will be specified for each exercise. I expect most students to meet these deadlines without difficulty. I shall accept late submissions until I begin returning corrected papers to other students. However, my policy is to judge late papers more strictly than ones that come in on time, on the theory that the extra time should be put to good use.

I prefer to receive your solutions in hard copy, printed by computer programs rather than handwritten. For preparing homework papers in hard copy (or Portable Document Format), I recommend the use of the \LaTeX document-processing utility. “The not so short introduction to \LaTeX,” by Tobias Oetiker, explains the basic use of this software. In addition, a useful users’ guide, “\LaTeX2e: An unofficial reference manual,” is available online. Some \LaTeX documentation refers to a “Local Guide” that describes how to run \LaTeX in the reader’s working environment. The third link below takes you to the Local Guide for MathLAN.


The tipa package supporting the fonts for the use of the International Phonetic Alphabet in \LaTeX documents, is available on MathLAN. Complete documentation can be found at


Class Attendance

I expect you to attend every session of the class and participate actively. It is especially helpful if you raise for discussion any questions you may have about the day’s topic, the assigned reading, or the exercises. I suggest that you write out such questions as part of your preparation for class sessions and pose them as opportunities arise.

Notwithstanding that general expectation, class attendance is optional (except for the final examination, which is obligatory). However, if you miss a class session for any reason, you must write up and submit solutions to certain designated exercises for that session.

Days of Religious Observance

Grinnell College acknowledges and embraces the religious diversity of its students, faculty, and staff. If you miss a class session in order to fulfill a religious obligation, your absence is excused. The accommodation is the one described in the preceding section: You must write up and submit solutions for the designated exercises for that session.

If one of your days of religious observance coincides with the final examination, please let me know early on in the semester so that we can arrange for you to take it at an earlier time.

In religious matters, I am a free thinker. I have no religious obligations that are tied to specific days, and I expect to attend all of the class sessions scheduled for this semester.
The Examination

The only examination in this course is a three-hour cumulative final examination scheduled for 9 a.m. on Thursday, December 15. It will be an anthology of short problems, proof constructions, and essay questions.

For the examination, you may bring and use any printed materials that you want to have available, but you may not use any electronic devices that are capable of computation, data processing, or communication.

Grading

Section IV.A.4 of the Faculty Handbook explains the duty of faculty members to report grades as follows:

Among the accepted responsibilities and obligations of each member of the faculty is that of reporting to the Registrar, at the appointed times, grades in accordance with the grading system and with the grading regulations which the faculty shall from time to time adopt. Further, faculty members are expected to make timely evaluations of students throughout the year.

Inherent in the responsibility of reporting grades is the further understanding that all such grades reported shall have been determined, in the final analysis, on the basis of the faculty members own professional evaluation of each individual students work.

This above indicated responsibility is considered to be part of the contractual relationship between the individual faculty member and the College, and the failure to fulfill this obligation will be considered a breach of contract.

In each of my courses, I fulfill this responsibility to the letter. However, I advise students that they should not regard the grades reported to the registrar as complete or even adequate assessments of their intellectual strengths and weaknesses. Grinnell College’s grading system is extremely inexpressive and vague.

The role of grades in the operation of the College is to support the generation of factoids (such as grade-point averages) that notionally provide quantitative justifications for decisions about progress towards graduation, off-campus study, academic honors, and such like. When collected on transcripts, grades also provide a convenient (though potentially misleading) way of summarizing academic performance for the benefit of personnel managers and graduate-school admission committees, who often cannot spare the time to read accurate assessments of candidates’ strengths and weaknesses.

I therefore urge you to pay much closer attention to your verbal interactions with me and to the comments that I make on your course work than to the grade that I report at the end of the course. Nonetheless, that grade is, as specified, based entirely on my professional evaluation of your individual work.

Your performance on exercises will determine seven-tenths of your final grade, class attendance and participation one-tenth, and your performance on the examination the remaining fifth, with the condition that you must pass the final examination in order to pass the course.

Copyrights and Licensing

Source code that I create for this course and put in the code repository on MathLAN at /home/stone/computational-linguistics/code is licensed under the GNU General Public License. Code released under the GPL can be copied, studied, revised, improved, and redistributed freely, subject to the restriction that any copies or derived programs are released under the same license.

Similarly, handouts and other prose course materials that I have created and distributed either in /home/stone/courses/computational-linguistics/handouts or in hard copy are licensed
under the Creative Commons Attribution–ShareAlike 4.0 International License and can be copied, studied, revised, improved, and redistributed freely, provided that my authorship of the original work is acknowledged and that any copies or derived works are released under the same license.

Under Grinnell College’s copyright policy, a student who submits a program or a prose text to satisfy a requirement of this course retains the copyright to it. Similarly, a group that submits such a work has and retains a collective copyright. In either case, however, the College asserts the right to distribute the work within the College community for instructional or administrative purposes without paying any royalty to the student.

You may, of course, choose to publish your work under some more generous license, and I encourage you to use the GNU General Public License (version 3) for software and the Creative Commons Attribution–ShareAlike 4.0 International License for text.

To my knowledge, the College has never abused its claimed right of distribution or used it in a way that would be contrary to the interests of student authors. In 2013, however, the College designated Microsoft Corporation as a school official with a “legitimate educational interest” in student records. Microsoft Corporation has a much sketchier track record that includes many instances of exploitation of its customers and users. Its rap sheet includes a felony conviction in American criminal courts on charges of abuse of monopoly and a similar conviction in the European Union. In addition, since 2007, Microsoft has given the National Security Agency access to most of its users’ emails, video chats, and cloud document storage, in unencrypted form.

In my opinion, Microsoft cannot be trusted to respect the rights and interests of student authors. Accordingly, I now advise students not to store any original writing, including source code for computer programs, on the College’s cloud-based storage servers and not to e-mail it (without encryption) to or from any @grinnell.edu address, since either of those acts might be construed as allowing Microsoft, as a school official, to read, copy, modify, distribute, and mine data from your original work without your permission.

Collaboration and Plagiarism

Since you will receive credit as an individual on the basis of your performance in this course, it would be unethical to submit a solution to any of the exercises that is not entirely your own work. To borrow other people’s solutions without acknowledgement is improper in any case; in this course, however, it is also improper to take answers or partial answers from others even if their contributions are explicitly acknowledged. Similarly, you may not collaborate on the examination or any part of it.

If I encounter clear indications of plagiarism or academic dishonesty, the Committee on Academic Standing will deal with them. The College’s policies for faculty members prohibit me from trying to investigate major offenses on my own. For the same reason, I impose penalties for academic dishonesty only as directed by the Committee on Academic Standing.

Because in recent years the Department of Computer Science has received mixed signals from the Committee on Academic Standing and from the Office of the Dean on the nature and scope of academic dishonesty, we have formulated departmental standards of academic integrity for students in our courses. They are available online at http://www.cs.grinnell.edu/academic-honesty-policy.