

CSCI 658
Natural Language Processing II
(updated 11/20/2023)

Fall 2023

TTh 2:00 - 3:15 PM

PM-110

1. Basic information

1.1. Instructor. Dr. Reva Freedman

1.2. Email address. rfreedman@niu.edu

1.3. Office hours. 3:30-5 PM Tue. and 12:30-2 PM Wed. in PM-554. If you would like to have a private conversation, I would be happy to make an appointment with you with a couple of days notice. I usually can't satisfy last-minute requests.

1.4. Course web sites. The course web site is <http://www.cs.niu.edu/~freedman/658/> . Most course materials will be on Blackboard under CSCI 658. Some sample NLP code will be on turing at ~t90rkf1/d658 . Some sample Python code will be on turing at ~t90rkf1/d503 .

1.5. Course goals. To gain an appreciation of the wide variety of natural language applications in computer science today, the difficulties involved in building such applications, and useful software and algorithms for doing so.

2. Communication

2.1. In person. I will be happy to talk to you about questions or concerns at any time, including any topic relating to this course, artificial intelligence or computer science in general. I am also happy to talk to you about personal or career issues to the extent that I can be helpful. I encourage you to address small problems before they become big problems, not the day before the exam.

2.2. Announcements. Late-breaking news, e.g., errors in assignments and weather emergencies, will be posted on Blackboard. Personal messages will be sent to your NIU email. I suggest you check both every morning.

2.3. University announcements. If the university is closed for weather or other reasons, class will automatically be cancelled. If weather or other emergencies arise, class cancellations will be announced as soon as possible via Blackboard.

3. Email

I receive more email than I can possibly answer, so I want to triage the email to be as helpful as possible to you. Therefore I prefer not to answer emails that you can answer from other sources.

3.1. Relevant email. I appreciate emails discussing bugs in assignments, suggestions for class, useful information, interesting articles, and similar things. I generally do not answer email if the

answer can be obtained elsewhere. If the answer to your question is in the syllabus, check the syllabus first. If you miss class, you are expected to find out what was covered from another student. If you have a reference question, e.g., the meaning of a specific error message, Google it first.

3.2. Email response schedule. I try to respond to email by the next business day. That means that an email you send on Tuesday should be answered no later than Wednesday. I also generally check my email once on Sunday in order to start the week without a backlog.

3.3. Questions about debugging. If you need debugging help, follow the directions on the course administration document as to the appropriate format to send. In general, you need to send something in the same format that you would submit. Include information on what happened and what you expected to happen. *Do not send pictures of text, i.e., copy and paste the text instead.* Questions requiring major debugging are usually best handled in person.

3.4. Do not send duplicate emails. There is no need to send email to more than one email address; my email addresses all forward to the same place. Don't send a second copy of an email if you haven't given me time to answer the first one.

4. Resources

4.1. No available textbook. Most of the NLP information will be provided on slides or online readings. Other possible references include the NLTK textbook, the Jurafsky and Martin textbook, the Manning and Schütze textbook and the Eisenstein textbook. Python information can be found in the slides, on the course web page and on the web in general.

4.2. Attendance and quizzes. I expect attendance since much of the material is not easily available elsewhere; however, I will not explicitly take attendance. There will occasionally be unannounced quizzes which will be extra credit. There are no makeup quizzes. If you arrive after the quiz, you have missed the quiz.

4.3. Class material. You are responsible for all material covered in class. If you miss a class, you must get notes from another student before the next class, *not from the instructor*. Research has shown that students who attend class regularly do better regardless of other behaviors.

4.4. Blackboard. Slides used in class will be posted at some point after the lectures in which they are used. Although many or most class materials will be posted on Blackboard, that is for your convenience. This is *not* an online class, and there is no assumption that you can learn everything without coming to class.

4.5. Class participation. Class participation is encouraged and will make the class more interesting for you and for other students. If you have a question, there are probably three other people with the same question who are even more shy than you.

Please ask ASAP if you don't understand, if I make a mistake, if you are wonder about the utility of an algorithm or its application, or if you are just curious about something. Questions about details, big ideas, concepts, algorithms, examples, related ideas and applications are all welcome.

5. Assignments

5.1. Types of assignments. Assignments may include experiments with existing software, larger programs and projects, and pencil-and-paper assignments. Some assignments will involve programming.

5.2. Due dates. In general, you will receive 10% extra points (rounded) for assignments submitted 24 hours before the deadline. There will be a penalty of 10% of the points for the assignment (rounded) for each day or portion of a day for late programs. No assignments will be accepted more than 2 days late. I suggest you submit your assignments by 11:58 PM to make sure they are received on time. If any assignments follow a different schedule, that will be specified on the assignment sheet.

In general, there are no waivers of the late penalty or extensions beyond this period, and no assignments may be submitted after the last day of classes.

5.3. Submission rules. All programs must run on the platform specified in the assignment. Code must be written in Python 3 unless otherwise specified. Regardless of how you develop your code, make sure you test it on the platform specified before you submit it.

5.4. Administrative penalties. There will be a penalty (generally 10 points) for not following directions, e.g., not following naming conventions, not following the class Python style guide, not submitting files in the correct format, not using the assigned directory structure, not submitting all the files, submitting extra files, and similar issues.

5.5. Naming conventions. Every student will be assigned a FLID (“four-letter ID”). Make sure that all quizzes, assignments and exams contain your 4-letter ID as well as the quiz, assignment or exam number.

Make sure that assignments submitted on Blackboard use the naming convention specified in the assignment, which includes your 4-letter ID and the assignment number. The naming convention is case-sensitive. For files, your 4-letter id is always lower case.

5.6. Grading criteria. Every assignment will include a rubric stating how points will be assigned. In addition, to receive full credit, programs must be comprehensible to humans. If multiple versions (e.g., source, object, and/or output) are required, they must be consistent.

5.7. Programming standards. Programs must follow standards on the course web page. You may not use external code (from other people or from the web) without permission, with the exception of code from the official Python libraries, numpy, pandas and other libraries specified in the assignment.

5.8. Grade correction deadline. If you believe your assignment has been graded incorrectly, you must see the instructor within one week after the grades have been posted.

6. Exams

6.1. Date of final exam. The final exam will be on the date scheduled by the registrar at <https://www.niu.edu/registration-records/dates/exams/index.shtml> under Standard Exams, namely Tuesday, Dec. 12, 2-3:15 PM. In exceptional cases, such as another Covid shutdown or a weather emergency, changes may be made to the exam schedule or format.

6.2. Form and content of exam. The exam will include material from the lecture notes and other class material. It will be a traditional closed-book, closed-notes exam and include mostly short essay questions.

6.3. Academic integrity policy for exams. Exams will be conducted in accordance with the department's academic integrity policy, which is available on the course web site.

6.4. Review sheets for exam. A review sheet will be posted with the topics to be covered and an outline of the possible questions.

6.5. No makeup exams (in general). You are expected to take the exam on the assigned time and date. Missing an exam is an extremely serious matter: makeup exams will only be given if *all* of the following requirements are satisfied: (a) an unavoidable reason (e.g., car accident), (b) advance notification, (c) written documentation, (d) permission of instructor. For a makeup final exam, all of the above are required in addition to permission of the department.

6.6. University final exam rule. Please notify me if you meet the university's criterion for rescheduling a final exam, namely that you have three finals on the same day and this course is the highest-numbered of the three. A makeup final can also be given if you have two overlapping finals. If the conflict results from another instructor changing the assigned time, that instructor is responsible for the makeup exam. The university deadline for scheduling a makeup final is Monday of the last week of classes. A printout of your schedule will be required.

7. Grade calculation

Each homework will be assigned a point count according to difficulty and the amount of time required.

There will be 3 assignments and one exam:

HW1 Text generation via n-grams.

HW2 Class presentation, or report on class presentations.

HW3 Machine learning using NLP input.

Final exam.

The final exam will count 25% of your grade. Each of the assignments will count approximately 25%. Unannounced quizzes will be extra credit.

Grades will not be curved or rounded. Grades will be calculated according to the following schedule:

Grade	Average
A	≥ 90
A-	≥ 89
B+	≥ 88
B	≥ 80
B-	≥ 79
C+	≥ 78
C	≥ 70
D	≥ 60
F	< 60

8. Classroom decorum

In classes for freshmen, I state the following rule on the syllabus: “No activity that interferes with learning, i.e., one that may distract other students or the instructor, is permitted in class. For example, eating, talking (whether in person or on the phone), newspaper reading, and regularly being late or leaving early are not permitted.” In a class for seniors and graduate students, I assume it is not necessary to state such a rule explicitly; however, the rule remains in force.

Research has shown that activities such as texting and checking your email interfere with *your* retention of the material, however, they in general do not interfere with other people’s ability to concentrate. Thus if you need to communicate with colleagues during class, texting and email are excellent ways to do so. Please do not sit in the front row if you intend to engage in these activities, as it is distracting.

9. Special circumstances

9.1. Types of special circumstances. Students with special needs (disability accommodation, religious observances, required military service, major illness or other unexpected events) are encouraged to contact the instructor as soon as possible. I am always willing to talk to students. There is a specific section on disability accommodations below.

9.2. Non-special circumstances. Having a lot of work for your other classes, being busy at your job, and network problems are *not* special circumstances; they are normal circumstances that everyone has.

10. Disability accommodations

The instructor will provide all of the accommodations to which you are entitled by law.

If you need an accommodation for this class, you must provide a notification letter from the Disability Resource Center. Once you provide a copy of the notification letter, we will have a private conference to determine how your approved accommodations will be handled in this class. This conference must be held and agreement reached before any accommodations can take effect. No accommodations will be allowed retroactively.

If you wish to take your exams at the DRC office, you must also follow DRC regulations with regard to exam scheduling.

For these reasons you should contact the DRC as soon as possible. They are located on the 4th floor of the Health Services Building, and can be reached at 815-753-1303 or drc@niu.edu. Also, please contact me privately as soon as possible to discuss possible accommodations – there is no need to wait until you have received the notification letter.

11. Academic integrity

You are encouraged to study together, however, that does not mean doing assignments together. Practice on problems from class, from the slides, or your own problems. Do the programs and any written assignments yourself.

You are expected to do your own work on the homework, programs and exams. Cheating includes, but is not limited to, copying work from other students, copying work from other textbooks, copying work from the Internet, or allowing others to do the same, whether deliberately or not.

You may not post material from this class, including answers to the homework assignments, on any public web site.

All cheating will result in the filing of an academic misconduct form and will affect your course grade, with the possibility of failing the course and/or losing your student job. The penalty for a first offense is usually two letter grades. Note that a second academic misconduct offense may result in your expulsion from the university.

We may use mechanized source comparison on the programs.