Natural Language Processing I  
*Fall 2020*

CSCI 657/490-B8  
TTh 2:00-3:15 PM  
online

**Instructor:** Dr. Reva Freedman  
**Email:** rfreedman@niu.edu  
**Office hours:** after class or by appointment  
**Course web site:** [http://faculty.cs.niu.edu/~freedman/657/](http://faculty.cs.niu.edu/~freedman/657/)  
Most course materials will be on Blackboard under CSCI 657  
Sample NLP code will be on turing at ~t90rkf1/d657  
Sample Python code will be on turing at ~t90rkf1/d503

**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.

**Communication with the professor:** I will be happy to talk to you about questions or concerns at any time, including any topic relating to this course or to artificial intelligence or computer science in general. I encourage you to address small problems before they become big problems, not the day before the exam.

I try to respond to email by the next business day, but there are occasional exceptions. Questions that can be answered from reference material may not be answered. Questions requiring major debugging are best handled in person.

If you need to send code, send a zip file using the same format as for submissions. Do not send pictures of any textual data, i.e., copy and paste the text instead.

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.

**Class schedule:** 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.

**Textbook:** There is one required textbook: *English Grammar: A Descriptive Linguistic Approach*, by Gülşat Aygen, third edition. There will be regular assignments from this 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.

**Software:** Software will come from some of the following or their equivalents: the Natural Language Toolkit (http://nltk.org), the Stanford NLP Group, Weka, numpy, scipy and scikit-learn.

**Attendance and quizzes:** Attendance is required. I will not explicitly take attendance. I will try to record the lectures. However, I occasionally give unannounced quizzes. There are no makeup
quizzes. If you arrive after the quiz, you have missed the quiz. The lowest two quiz scores will be dropped. I do not warrant that you can follow the class from the recordings.

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.

Although many or most class materials will be posted on Blackboard, that is for your convenience. This is not an asynchronous class, and there is no assumption that you can learn everything without coming to class.

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.

Assignments: Assignments may include experiments with existing software, larger programs and projects, and pencil-and-paper assignments. Graduate students will be required to do an additional assignment or a more advanced version of one or more assignments. Graduate students may be required to do an oral presentation on their work.

Unless otherwise specified, the following rules apply to all assignments. Assignments must be submitted as defined in the assignment writeup. You will receive a bonus for assignments submitted 24 hours before the deadline. Late homework is not accepted. In general, there are no waivers or extensions. No assignments may be submitted after the last day of classes. I suggest you submit your assignments by 11:57 PM to make sure they are received on time. Exceptions to this rule may be made for specific assignments.

Your 4 letter ID is the first four characters of your last name as it is recorded on myNIU, in lower case. If your last name is shorter than 4 letters, fill it out with 0’s, e.g., oh -> oh00. 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.

To receive full credit, programs must (a) work, (b) follow the specifications, (c) be comprehensible to humans, (d) be accompanied by any requested writeup. If multiple versions (e.g., source, object, and/or output) are required, they must be consistent.

There will be a penalty if your program does not follow the coding, layout, documentation or submission standards referred to in the assignment, for example, e.g., if you do not submit the correct format, do not submit all the files, submit extra files or submit files with incorrect names.

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. All programs must run in Python 3 on turing/hopper unless otherwise specified. I encourage you to develop them on turing, but if you don’t, make sure you test them on turing before you submit them. There will be a penalty for not following directions, e.g., not following naming conventions, not following or the class Python style guide, not submitting
files in the correct format, not using the assigned directory structure, not submitting all the files or submitting extra files.

If you believe your assignment has been graded incorrectly, you must see the instructor within one week after the assignments have been returned.

**Exams:** There will be a midterm exam and a final. Dates of the midterm will be announced approximately one week in advance. The final exam will take place during the final exam slot. The exams will most likely be traditional closed-book, closed-notes exams, but could also be a take-home exam or a project presentation. Exams will be conducted in accordance with the department’s academic integrity policy, which is available on the course web site.

With regard to concepts, each exam will cover one section of the course material (the final will not be cumulative). With regard to programming, programming is inherently a cumulative activity, so the exams may include programming constructs from earlier in the course.

Exams will include material from the lecture notes, written assignments and programs. For each exam, a review sheet will be posted listing all the possible conceptual questions (in a somewhat different format). Exam formats will include multiple choice, fill-in-the-blank and similar formats. They are also likely to include some short functions to write (10-15 lines or so). You will not be required to memorize names of libraries or similar issues. There may be some short essay questions. Slides used in class will be posted at some point after the lectures in which they are used.

Examples for the programming questions on the exams will be available from sample code and the homework. Programming questions will assume that you have not only done the homework but learned from it, i.e., copying code from the sample programs used in class without understanding them may give you a working program but is less likely to give you the level of understanding you will need for the exams.

You are expected to take the exams 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, (e) for final exam, permission of department.

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. The university deadline for scheduling a makeup final is Monday of the last week of classes. Documentation will be required.

**Grading:** Each homework will be assigned a point count according to difficulty and the amount of time required. Grades will be calculated as follows: Exams 40%, assignments 50%, quizzes 10% (grades will not be curved or rounded).
<table>
<thead>
<tr>
<th>Grade</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>&gt;= 90</td>
</tr>
<tr>
<td>A-</td>
<td>&gt;= 89</td>
</tr>
<tr>
<td>B+</td>
<td>&gt;= 88</td>
</tr>
<tr>
<td>B</td>
<td>&gt;= 80</td>
</tr>
<tr>
<td>C+</td>
<td>&gt;= 78</td>
</tr>
<tr>
<td>C</td>
<td>&gt;= 70</td>
</tr>
<tr>
<td>D</td>
<td>&gt;= 60</td>
</tr>
<tr>
<td>F</td>
<td>&lt; 60</td>
</tr>
</tbody>
</table>

Class 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. Therefore these are permitted in case of emergency. Please do not sit in the front row if you intend to engage in these activities, as it is distracting.

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. 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.

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.

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.