CSCE 4290/5290: Natural Language Processing (Online)

Spring 2024

Course Information & Syllabus

**Instructor:** Dr. Zeenat Tariq

**Pronouns:** she/her

**Office:** E235M Discovery Park

**Office timings:** T/Th 4:30 PM -5:30 PM (by appointment)

**Email:** zeenat.tariq@unt.edu

**Course details:** This course will be instructed in an asynchronous mode. In addition to asynchronous learning, I understand the importance of interactive sessions for a comprehensive understanding of the material. Therefore, we will be conducting live and recorded tutorial sessions for practicals. These sessions aim to enhance your learning experience by providing opportunities for real-time engagement and clarification of concepts.

A detailed schedule for the online tutorial sessions will be announced in advance, allowing you to plan your participation accordingly. We encourage you to take advantage of both the asynchronous and live/recorded sessions to make the most out of your learning journey.

**Course Description**

Natural language processing (NLP) is a subfield of AI focused on solving problems that involve dealing with human language in a sophisticated way: these include information extraction, machine translation, automatic summarization, conversational dialogue, syntactic analysis, and many others. Much of the progress on these problems over the last 25 years has been driven by statistical machine learning and, more recently, deep learning. One distinctive feature of language compared to other types of data is its structured nature: modeling language involves understanding the linguistic phenomena it exhibits and grappling with it as a sequentially structured, tree-structured, or graph-structured entity. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages.

**Course Objectives**
This course introduces the field of computational linguistics, aka natural language processing (NLP). We will learn how to create systems that can understand and produce language, for applications such as information extraction, machine translation, automatic summarization, question-answering, and sentiment analysis. The course will cover linguistic (knowledge-based) and statistical approaches to language processing in the three major subfields of NLP: syntax (language structures), semantics (language meaning), and pragmatics/discourse (the interpretation of language in context). Homework assignments will reflect research problems computational linguists currently work on, including analyzing and extracting information from large online corpora.

In addition to “lecture” classes, each student will participate in an in-class coding exercise if required.

**Learning Outcomes:**

By the end of the course, students will be able to

1. Extract information from text automatically using concepts and methods from natural language processing (NLP) including stemming, n-grams, POS tagging, and parsing.
2. Analyze the syntax, semantics, and pragmatics of a statement written in a natural language.
3. Write scripts and applications in Python to carry out natural language processing using libraries such as NLTK, and spaCy.
5. Apply basic machine learning algorithms to natural language processing.
6. Evaluate the performance of NLP tools and systems.

**Course Prerequisites**

- Familiarity with Programming (for programming assignments and activities). Understanding of basic python language is required.
- Basic understanding of linguistics, mathematics and statistics i.e., probability, linear algebra, and optimization is required.
- Machine learning, Artificial Intelligence, or equivalent AI/ML course experience.

**Targeted audience:** Graduate and undergraduate students from Computer Science and related areas.

**Recommended Textbooks**

“Speech and Language Processing, by Jurafsky and Martin”

“Natural Language Processing, Jacob Einstein”
Additionally the lectures and topic-based tutorials will be posted via Canvas throughout the semester.

**Tentative Topics and Course Schedule (subject to change based on learning rate in class)**

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Course/Topics</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 15 - 20</td>
<td>Course Overview, Introduction</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Jan 21 - 27</td>
<td>Regular Expressions</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Jan 28 - Feb 3</td>
<td>Basic Text Processing, Normalization</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Feb 4 - 10</td>
<td>Minimum Edit Distance</td>
<td>Project idea/Proposal</td>
</tr>
<tr>
<td>5</td>
<td>Feb 11 - 17</td>
<td>Language Modeling</td>
<td>Quiz 1</td>
</tr>
<tr>
<td>6</td>
<td>Feb 18 - 24</td>
<td>Language Modeling (Smoothing Techniques)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Feb 25 - Mar 2</td>
<td>Spelling correction</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>8</td>
<td>Mar 3 - 9</td>
<td>Text Classification</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Mar 10 - 16</td>
<td>Spring Break</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Mar 17 - 23</td>
<td>Machine learning</td>
<td>Quiz2</td>
</tr>
<tr>
<td>11</td>
<td>Mar 24 - 30</td>
<td>Sentiment Analysis</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Apr 7 - Apr 13</td>
<td>Question Answering</td>
<td>Assignment 2</td>
</tr>
<tr>
<td>13</td>
<td>Apr 14 - Apr 20</td>
<td>Text Summarization</td>
<td>Quiz 3</td>
</tr>
<tr>
<td>14</td>
<td>Apr 21 - 27</td>
<td>PoS Tagging</td>
<td></td>
</tr>
</tbody>
</table>
Teaching Philosophy

I believe that the best way to really learn and understand concepts in computer science is to implement the techniques and perform the operations in an applied manner. While the process of reading examples and proofs in textbooks and lecture notes is valuable, real learning comes through one’s own efforts in debugging and solving logical and programming problems. I have a very hands-on approach to teaching, which includes coding-on-the-fly and describing how programs work, which typically requires students to be present in the lecture. I would encourage students to use this opportunity to really learn and develop the skills covered in this class that will prepare you for entering a career in industry or furthering your education.

Course Requirements

Quizzes, Participation activities (hands on experience), Assignment, Conceptual Project Presentation, Exam

Participation activities: Activities/Tutorials are designed to engage you in your learning, so you can begin to apply these principles in practice and tailor them to your needs. You will get hands on different concepts/examples in NLP. Your participation in activities will be counted towards your final grade. There will be no specific weekly due dates for the activities. The activities will be given during the online session(weekly/bi weekly). However, if some students are unable to attend the session they can complete it by going over the recorded videos.

Assignments: There will be 2 assignments in the semester. The due date will be specified once the assignment is posted. Reports are to be turned in as PDF. Code is to be turned in with both Jupyter notebook/source code and PDF form, along with any files necessary to run your assignment. Results should be presentable, with appropriate comments for someone to follow what you have done. All assignments must be turned in individually, although students are encouraged to work together extensively.

Project Presentation (Conceptual): After a few weeks into the course, you will be working on project (conceptual idea) and present the work by the end of semester. You are required to work as a team. However, due to less interaction you will be allowed to work on the project individually. This is your opportunity to demonstrate what you have learned in a way that
reaches beyond the selection of tools, data sets, and approaches demonstrated in the course. Commonly students find a unique, complex data set and associated learning problem and apply the techniques presented in the class. The goal here is to create a coherent, completed work for presentation at the end of class. Essentially ask yourself what you would want to show an employer (or brag about to others) demonstrating what you have learned in the course.

Quizzes: These quizzes are meant to focus students on the important aspects of the readings or lectures. You will be taking the quizzes on canvas using lockdown browser. You will be notified about the quiz one week prior taking it.

Exam: There will be one final Exam taken on canvas using lockdown browser. Exam cannot be missed without prior arrangements or later documented proof of extenuating circumstances.

Grading

Grades are determined by a simple points system, with a total of at least 100 pts as the goal though more than 100 points are likely. The expected distribution of points is given below, with the exact scale determined by point values given for each component - this is subject to minor modification based on actual points given.

- 25% Participation activities
- 20% Assignments (2)
- 25% Presentation (Project Presentation)
- 15% Quizzes (3)
- 15% Exam
- Bonus (Possible one extra credit activity)

Grading Scale: A=90, B=80-89, C=70-79, D=60-69, F=0-59 pts. No exceptions. If class grades are low (e.g., I expect most students will end with A’s and B’s), rescaling of some of the components or extra assignment will be given to add points to the class.

CSCE 4290 vs 5290 grading: In line with the added expectations for the graduate version of the course, students enrolled in CSCE 4290 will receive 5 additional bonus points.

Course Technology & Skills

Strongly Recommended: Laptop/computer for practical work.

UNT Policies

Academic Integrity Policy
Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

ADA Policy

UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one’s specific course needs. Students may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website (https://disability.unt.edu/).

Emergency Notification & Procedures

UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

Retention of Student Records

Student records pertaining to this course are maintained in a secure location by the instructor of record. All records such as exams, answer sheets (with keys), and written papers submitted during the duration of the course are kept for at least one calendar year after course completion. Course work completed via the Blackboard online system, including grading information and comments, is also stored in a safe electronic environment for one year. Students have the right to view their individual record; however, information about student’s records will not be divulged to other individuals without proper written consent. Students are encouraged to review the Public Information Policy and the Family Educational Rights and Privacy Act (FERPA) laws and the University’s policy. See UNT Policy 10.10, Records Management and Retention for additional information.

Acceptable Student Behavior

Student behavior that interferes with an instructor’s ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom,
labs, discussion groups, field trips, etc. Visit UNT’s [Code of Student Conduct](https://deanofstudents.unt.edu/conduct) to learn more.

**Access to Information - Eagle Connect**

Students’ access point for business and academic services at UNT is located at: [my.unt.edu](http://my.unt.edu). All official communication from the University will be delivered to a student’s Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail [Eagle Connect](https://it.unt.edu/eagleconnect).

**Student Evaluation Administration Dates**

Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via IASystem Notification" (no-reply@iasystem.org) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the [SPOT website](http://spot.unt.edu/) or email spot@unt.edu.

**Getting Help**

**Technical Assistance**

[UIT Help Desk](http://www.unt.edu/helpdesk/index.htm)

**Email:** helpdesk@unt.edu

**Phone:** 940-565-2324

**In Person:** Sage Hall, Room 130

**Walk-In Availability:** 8am-9pm

**Telephone Availability:**

- Sunday: noon-midnight
- Monday-Thursday: 8am-midnight
- Friday: 8am-8pm
- Saturday: 9am-5pm

**Laptop Checkout:** 8am-7pm
Student Support Services

- Registrar (https://registrar.unt.edu/registration)
- Financial Aid (https://financialaid.unt.edu/)
- Student Legal Services (https://studentaffairs.unt.edu/student-legal-services)
- Career Center (https://studentaffairs.unt.edu/career-center)
- Multicultural Center (https://edo.unt.edu/multicultural-center)
- Counseling and Testing Services (https://studentaffairs.unt.edu/counseling-and-testing-services)
- Student Affairs Care Team (https://studentaffairs.unt.edu/care)
- Student Health and Wellness Center (https://studentaffairs.unt.edu/student-health-and-wellness-center)
- Pride Alliance (https://edo.unt.edu/pridealliance)

Academic Support Services

- Academic Resource Center (https://clear.unt.edu/canvas/student-resources)
- Academic Success Center (https://success.unt.edu/asc)
- UNT Libraries (https://library.unt.edu/)
- Writing Lab (http://writingcenter.unt.edu/)
- MathLab (https://math.unt.edu/mathlab)