

**Natural Language Processing
CAP 4641**

Section: 1ZED, 6ZYX [FTF and Virtual]

Class Periods: T Periods 8-10 (3:00 PM – 6:00 PM)

Location: [MCCB G086](#) and Zoom [see weekly Modules in Canvas]

Academic Term: Fall, 2024

Instructor:

Bonnie J Dorr

bonniejdorr@ufl.edu

352-273-4045

Office Hours: By appointment (request 24 hours in advance), as specified in Canvas

Teaching Assistants and Mentors:

Please contact through Canvas website

Shlok Gilda (TA), shlokgilda@ufl.edu

Chathuri Jayaweera (TA), chathuri.jayawee@ufl.edu

Sangpil Youm (TA), youms@ufl.edu

Amanda Smith (Mentor), a.smith3@ufl.edu

Bryan Etzine (Mentor), bryanetzine@ufl.edu

Course Description

Introduction to the essential concepts, principles, and techniques of Natural Language Processing (NLP). Practical application and theoretical concepts are examined. Topics include information extraction, language construction, grammars, disambiguation, as well as system modeling, classification, and evaluation.

Course Pre-Requisites / Co-Requisites

Prerequisite: COP 3530

Course Objectives

Students will learn the theory and practice of Natural Language Processing. Among the theoretical topics included are: parts of speech, n-gram models, context-free grammars, parsing and ambiguity, dependency parsing, sequence labeling, hidden markov models, vector semantics and embeddings, and semantic role labeling. Practically, we are motivated to study natural language to better understand how to develop a range of applications, including machine translation systems, conversational agents, and techniques for evaluating NLP output. The course also covers the application of machine learning solutions to natural language data. **Python** will be introduced to illustrate course concepts. Project solutions will be developed in Python. Students will be able to apply their knowledge to:

- implement and evaluate their own programmatic solutions to natural language problems
- apply symbolic and non-symbolic techniques to natural language problems
- evaluate the performance of natural language applications

Materials and Supply Fees There are no materials or supply fees for this course.

Relation to Program Outcomes (ABET):

Outcome	Coverage*
1. An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics	High

2. An ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors	High
3. An ability to communicate effectively with a range of audiences	
4. An ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts	Medium
5. An ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives	
6. An ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions	
7. An ability to acquire and apply new knowledge as needed, using appropriate learning strategies	High

*Coverage is given as high, medium, or low. An empty box indicates that this outcome is not covered or assessed in the course.

Required Textbooks and Software

All readings are online and freely available:

- The third edition of *Speech and Language Processing* is published here: <https://web.stanford.edu/~jurafsky/slp3/>
- The second edition of *Natural Language Processing with Python* is published here: <http://www.nltk.org/book/>
- A primer on spaCy tool for NLP is published here: <https://course.spacy.io/en>

Supplemental/Recommended Materials

Recommended online primers, tutorials, resources, and other texts may include:

- Python: <http://www.python.org/>
- Python refresher: <https://wiki.python.org/moin/BeginnersGuide>
- Learn and Practice Python: <http://inventwithpython.com/>

Required Computer

UF student computing requirement: <https://news.it.ufl.edu/education/student-computing-requirements-for-uf/>

Course Schedule

Here is a general schedule for the semester. Please refer to the Canvas *Announcements* and *Modules* for final details about each topic, exact dates, and the specific schedule we implement this semester.

Theme/Weeks	Weeks	Topics	Readings, Assessments due
Introduction	Weeks 0-1 8/22, 8/27	What is NLP? What are LLMs?	Wk0 Slides/videos NLP and LLMs, Review syllabus Quiz 0: due Wk1 (Thu 8/29, 11:59pm)
		Motivation, Applications, NLP Pipeline and Ambiguity	Wk1 Slides/videos SLP, 2.1 (parts), 2.2 Quiz 1: due Wk 1 (Thu 8/29, 11:59pm)

Words and Corpora	Weeks 2-3 9/3, 9/10	Word-level analysis of Corpora, Text Processing in Python	Wk2 Slides/videos SLP 2.3-2.7, spaCy 101; NLTK; AG's News Articles Quiz 2: also due Wk2 (Tue 9/3, 3pm) Proj 0: out 9/3, 6pm
		N-gram Language Models	Wk3 Slides/videos SLP 3.1, 3.2-3.5 (parts) Quiz 3: due Wk3 (Tue 9/10, 3pm) Proj 0: due Thu 9/12, 11:59pm
Linguistic Structure	Weeks 4-5 9/17, 9/24	Constituency Grammars, Parsing	Wk4 Slides/videos SLP 17.1-2, 17.5 Quiz 4: due Wk 4 (Tue 9/17, 3pm)
		Ambiguity, CKY, Dependency Parsing	Wk5 Slides/videos SLP 17.4, 17.6, 18.1-18.2 Universal Dependencies Quiz 5: due Wk 5 (Tue 9/24, 3pm)
Sequence Labeling	Weeks 6-7 10/1, 10/8	Sequence Labeling for POS and NE	Wk6 Slides/videos SLP, 8.1-8.3, 17.3 LDC Penn Treebank Quiz 6: due Wk6 (Tue 10/1, 3pm) Proj 1: out 10/1, 6pm SLP (ed 2) 13.5 "Chunking" (in Canvas)
		Information Extraction, Chunking, HMMs, Viterbi, Conditional Random Fields; Python NLP/ML Tools	Wk7 Slides/videos NLTK 0, 2, 7; SLP 8.4-8.5 Scikit Learn, CRFSuite Quiz 7: due Wk7 (Tue 10/8, 3pm) Proj 1: due Thu 10/10, 11:59pm
Semantics I	Weeks 8-9 10/15, 10/22	Word Meaning, Vector Semantics, and Embeddings, Review for Exam 1	Wk8 Slides/videos SLP, 6.1, 6.2, 6.3 Quiz 8: due Wk8 (Tue 10/15, 3pm) Proj 2: out 10/15, 6pm Exam 1: Tue 10/15 6pm-Wed 10/16 10pm
		Word Similarity: Cosine, Word2vec, TFIDF, Jaccard, Intro to Metrics: Precision, Recall, F-Measure	Wk9 Slides/videos SLP, 6.4, 6.5, 6.8, Slor, ElasticSearch; A Simple Explanation of the Jaccard Similarity Index, SLP 4.7 Quiz 9: due Wk9 (Tue 10/22, 3pm)
Semantics II	Weeks 10-11 10/29, 11/5	Semantic Role Labeling and multilingual projection; applications research talk TBD.	SLP 24 Quiz 10: due Wk10 (Tue 10/29, 3pm) Proj 2: due Thu 10/31, 11:59pm Proj 3: out 10/29, 6pm
		Contextualized embeddings; applications research talk TBD.	SLP, 11.1, 11.2, 11.3 (parts) Quiz 11: due Wk11 (Tue 11/5, 3pm)

Applications, Eval <i>[UF has no classes the week of Tgiving, no class or assignments on 11/26]</i>	Weeks 12-13 11/12, 11/19	Machine translation, Divergences, Encoder-Decoders, Start metrics (Bleu, ROUGE)	13.1-13.3 (parts), 10.2, 10.4(parts), additional notes
		Review Quizzes 10, 11; Finish metrics; research talk TBD.	SLP 13.6.1-13.6.2 and additional notes Quiz 12: due Wk13 (Tue 11/19, 3pm) Proj 3: due Thu 11/21, 11:59pm
Review/Last Day	Week 14 12/3 [last day]	Review	Exam 2: Tue 12/3 6pm - Wed 12/4 10pm

Important Dates:

- [Thu, 11:59pm](#) First two quizzes due: 8/29 (Q0/Q1)
- [Tue, 3pm](#) Remaining 11 Quizzes due: 9/3 (Q2), 9/10 (Q3), 9/17 (Q4), 9/24 (Q5), 10/1 (Q6), 10/8 (Q7), 10/15 (Q8), 10/22 (Q9), 10/29 (Q10), 11/5 (Q11), 11/19 (Q12).
- [Thu, 11:59pm](#) Projects due: 9/12 (Proj 0), 10/10 (Proj 1), 11/7 (Proj 2), 11/21 (Proj 3).
- [Tue 6pm-Wed 10pm](#) Exams due: 10/15-10/16 (Exam 1), 12/3-12/4 (Exam 2)

Attendance Policy, Class Expectations, and Makeup Policy

Class time occurs *synchronously* in our assigned face-to-face classroom [F2F] and virtually [online] via **Zoom: Tues Periods 8-10 (3:00 PM – 6:00 PM EST)**.

Format: This course is taught as a “flipped classroom.” All lectures are pre-recorded and delivered for viewing outside of the classroom in the form of 3-5 brief (5 to 15 min) video snippets per week. These videos are accompanied by a handful of practice questions (“quizzes”) to solidify understanding of the lecture material. You are responsible for watching all video snippets (and taking all quizzes) by the next class period **at 3pm each Tuesday**, starting with those released on the first day of the semester.

Class time will also be recorded and is dedicated to: (a) discussion of the current module, including review of practice questions (“quizzes”) and discussion of module-relevant “thought questions”; (b) problem solving based on students’ questions about current quizzes, projects, and exams; (c) discussion of the upcoming module via additional “thought questions.” To be successful in this course, you are advised to attend class sessions (either F2F or virtually) and to take advantage of this unique flipped classroom opportunity. There is no replacement for the interactive experience of discussing course material, problem solving, reviewing prior modules and preparing for upcoming modules.

General Attendance: Please pay attention and do not talk to fellow students during class. If you have a comment about another student’s question, please raise your hand (either in person or using Zoom technology) and tell everyone your observations or follow-up question. If you miss a class *never* email the professor or TA asking what was covered. Check inside Canvas for posted videos and notes and/or ask fellow students for their notes on that day’s class period. Here are additional points of significance:

- You are responsible for all material covered during pre-recorded lectures and class discussions.
- Class attendance will not be taken (in-person attendance is not mandatory).
- The 3-hour class time will be recorded (see the disclaimers and notice that follow).

Disclaimer #1: Zoom sessions are being held campus-wide to deliver courses and the demand for recorded materials is high. Thus, there may be delays completing the recording process. *I will post live*

classroom recordings as soon possible [under the corresponding Canvas Module], however the way to ensure you see the classroom activities is to participate in the synchronous FTF classroom sessions.

Disclaimer #2: The Zoom platform records these sessions and I cannot guarantee they will not experience errors in the recording process. Please be aware that attending the live classroom sessions is always the best way to ensure you are able to view the complete content of the discussion.

Notice: Our classroom discussion is audio visually recorded for students in the course to reference after the live recorded session. Students who choose to participate with their camera engaged or utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated. A text “chat” feature is also available during lecture. “Chats” sent through the public channel may be recorded. If you do not wish for your “chat” to be recorded, you must send it through the private channel directly and only to the professor. Since lecture is a public setting, private chats may or may not always be answerable during the live lecture. As in all courses, unauthorized recording and unauthorized sharing of recorded materials is prohibited.

Excused absences: There are very few events that would impede your participation in a graded activity warranting an exception, that you would not be aware of before the start and throughout the multi-day duration of that activity (quizzes, projects, and exams). The Dean of Students Office (www.dso.ufl.edu) provides rules, guidance, and approval for excuse documentation. The instructor notification letter from the DSO must directly state that an “**absence has been excused**” and the letter must specify the dates of the entire multi-day duration of the assessment. If the letter does not state this, the student must provide the instructor with documentation that meets UF’s rules for an excused absence (<https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/#absencestext>). Otherwise, the absence is not excused, and accommodations cannot be provided.

Makeups or extensions: Exam makeups or extensions are provided if appropriate documentation is presented in a timely fashion (see above). However, flexibility is already built into the flipped classroom schedule, where exams are conducted in a 2-3 hour time period that you select within a 1.5 day window. This generally avoids the need for exam makeups or extensions and allows you to plan ahead for a timeframe that easily accommodates scheduling constraints.

Late Assignments: All quizzes are due at 3pm the day of the lecture. There are absolutely no exceptions. We discuss solutions in class, so if the deadline is missed, the score for that quiz is zero. However, the bottom 4 quiz scores (out of 12 quizzes) are dropped for computing the final grade. So if you are unable to finish a quiz in time because of an emergency, you still have lots of chances (8 quizzes) to contribute toward your quiz total. There is no makeup due to this already very generous policy. **For projects and exams, a 50% penalty will be applied** if your submission arrives after the deadline. Any project or exam that is submitted **more than 24 hours after the deadline cannot be accepted**. For example, a project that is due on Thursday at 11:59pm is **late** starting after 11:59pm Thursday (even if just one second late), and will be graded 50% off up until 11:59pm on Friday. No credit is awarded starting after 11:59pm Friday (even one second later).

Questions outside of class time: A zoom meeting with the professor or TA can be scheduled for private discussions for matters that cannot be handled in a public classroom setting, e.g., to discuss a concern regarding a grade. Detailed technical questions are best addressed within the flipped classroom. If you raise a question outside of classroom time that is better addressed in the flipped classroom setting, the professor/TA may ask you to raise the question to the whole class. The 2.5 hours

of interactive FTF/virtual meeting time each week are intended to serve as a more effective and open setting for problem solving, discussion of class themes, preparing for upcoming assessments, and reviewing key elements of graded activities. It is advisable to show up each week with questions in hand, so that these are answered during classroom time and your classmates can also benefit from that discussion. This reduces redundancy as many students have an interest in the same questions. Please consider taking advantage of this unique flipped classroom experience before contacting the professor or TA to meet outside of the classroom for questions pertaining quizzes, projects, and exams.

Evaluation of Grades

Assessments are scheduled in non-overlapping multi-day time windows. The “flipped classroom” format provides ample time for the entire class to prepare interactively (together) for exams and projects, and to review the highlights of graded submissions (including weekly quizzes).

Activity	Total Points	Final Grade Percentage
<i>Total Quizzes (with videos)</i>	<i>32 (drop four bottom quizzes)</i>	20%
<i>Projects:</i>		
• <i>Project 0</i>	15	7%
• <i>Project 1</i>	20	9%
• <i>Project 2</i>	27	13%
• <i>Project 3</i>	35	16%
<i>Exam 1</i>	27	16%
<i>Exam 2</i>	33	19%
Total	189	100%

Videos and Quizzes: The “flipped classroom” design is conducted via modules with video snippets, coupled with quizzes in Canvas *Quiz* format. Quizzes highlight key video points to aid students in preparing for projects and exams. Quizzes are open book, with **one quiz attempt** allowed. **No late submissions are accepted.** Thus, it is *highly* recommended that you watch the videos, and download and review the associated PDF slides (and video transcripts) before taking the quiz. Quizzes must be completed **before** you attend the next class on Tuesday at 3pm, where solutions will be discussed. Of the 12 quizzes (each worth 4 points), the lowest four are dropped, with the remaining eight (a total of 32 points) counting for **20% of your final grade.**

Exams: There will be two exams, both electronically administered using Canvas *Quizzes*. Exams are open book, with access to any resource made available during the course, and will be held during a 120-180 min 28 hour time period that you select, between Tuesday after class and Thursday night (at 10pm). **Please plan ahead.** The exams count for **35% of your final grade.**

Projects: There are four projects, also electronically administered. Projects involve programming in Python using a Jupyter notebook format (e.g., Google Colab), coupled with both structured (automatically graded) and unstructured (human-inspected) responses to questions about central aspects of the programming assignment. Access to resources specified in projects is strongly encouraged. The details of project submissions will be given within each project specification. Projects are generally assigned on Tuesday after class and are due Thursday night (at 11:59pm) of the following week. **Please start your projects early.** The projects count for **45% of your final grade.**

Grading Policy

The range used to calculate your final letter grade in our course will be no harsher than the grading scale provided in this table. Your final point total will be calculated using the percentages given in the *Evaluation of Grades* section. The percent you earn on each activity will be multiplied by the grade points associated with that activity. More information on UF grading policy may be found at: <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>

Re-Grades

All assessments are electronic and automatically graded, with a complete record of all attempts/saves, final answers, and the automatic grade that has been assigned. If you believe there is an error in a question or its possible answer or you notice a technology issue, it is important to bring this to the attention of the professor or TA so that it can be addressed. A question can be raised in class once an assessment is completed and discussed openly in the class period. The professor/TA will address this in a timely fashion. Re-grades typically impact the entire class. If an adjustment is made, it will be applied electronically in one fell swoop for the entire class.

Right to change information

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester may require adjustments. Consequently, given due notice to students, the instructor reserves the right to change any information on this syllabus or in other course materials.

Grade Points	Letter Grade	Highest	Lowest
4.00	A	100+	92.00
3.67	A-	91.99	89.00
3.33	B+	88.99	86.00
3.00	B	85.99	82.00
2.67	B-	81.99	79.00
2.33	C+	78.99	76.00
2.00	C	75.99	72.00
1.67	C-	71.99	69.00
1.33	D+	68.99	66.00
1.00	D	65.99	62.00
0.67	D-	61.99	59.00
0.00	E	58.99	0.00

More information on UF grading policy may be found at:

<https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx>

Students Requiring Accommodations

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting <https://disability.ufl.edu/students/get-started/>. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback

in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

In-Class Recording

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

University Honesty Policy

UF students are bound by The Honor Pledge which states, “We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: “On my honor, I have neither given nor received unauthorized aid in doing this assignment.” The Honor Code (<https://sccr.dso.ufl.edu/process/student-conduct-code/>) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment

The Herbert Wertheim College of Engineering values varied perspectives and lived experiences within our community and is committed to supporting the University’s core values, including the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information, and veteran status.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:

- Your academic advisor or Undergraduate Coordinator
- HWCOE Human Resources, 352-392-0904, student-support-hr@eng.ufl.edu
- Pam Dickrell, Associate Dean of Student Affairs, 352-392-2177, pld@ufl.edu
- Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: <https://registrar.ufl.edu/ferpa.html>

Campus Resources:

Health and Wellness

U Matter, We Care:

Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: <https://counseling.ufl.edu>, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence

If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the **Office of Title IX Compliance**, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)

Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or <http://www.police.ufl.edu/>.

Academic Resources

E-learning technical support, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. <https://elearning.ufl.edu/>.

Career Connections Center, Reitz Union, 392-1601. Career assistance and counseling; <https://career.ufl.edu>.

Library Support, <http://cms.uflib.ufl.edu/ask>. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring. <https://teachingcenter.ufl.edu/>.

Writing Studio, 302 Tigert Hall, 846-1138. Help brainstorming, formatting, and writing papers. <https://writing.ufl.edu/writing-studio/>.

Student Complaints Campus: <https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/>; <https://care.dso.ufl.edu>.

On-Line Students Complaints: <https://distance.ufl.edu/getting-help/>;
<https://distance.ufl.edu/state-authorization-status/#student-complaint>.