

**Natural Language Processing  
CAP 4641**

**Section:** 5ZED [Hybrid: 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, 2022

**Instructor:**

Bonnie J Dorr

[bonniejorr@ufl.edu](mailto:bonniejorr@ufl.edu)

352-273-4045

Office Hours: By appointment (request 24 hours in advance), zoom link distributed within Canvas

**Teaching Assistant:**

Please contact through Canvas website

Chathuri Jayaweera, [chathuri.jayawee@ufl.edu](mailto:chathuri.jayawee@ufl.edu), zoom link distributed within Canvas

**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.

**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 Textbooks**

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/>

## Resources

Other suggested resources are:

- NLP on Wikipedia: [http://en.wikipedia.org/wiki/Natural\\_language\\_processing](http://en.wikipedia.org/wiki/Natural_language_processing).
- Association for the Advancement of Artificial Intelligence (AAAI): <http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/NaturalLanguage>.

## 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.

<b>CAP 4641: Tues 3-6pm ET, starting Aug 30, 2022. First day of semester is Wed Aug 24 (intro slides/videos go live on this date). Last class: Dec 6. Holidays: Sept 5, Nov 11, Nov 23-25.</b>			
<b>Theme/Weeks</b>	<b>Class Dates</b>	<b>Topics</b>	<b>Readings</b>
<b>Introduction 1-2: Aug 24-Sep 2</b>	8/30	What is NLP?	Intro slides/videos
		NLP Pipeline and Text Pre-processing	SLP, 2.2, 2.3, 2.4
<b>Words and Corpora 3-4: Sep 6-Sep 16</b>	9/6, 9/13	N-gram Language Models	SLP, 3.1
		NLP and Information Extraction in Python	spaCy 101
		Labeled Corpus Example 1	AG's News Articles
<b>Linguistic Structure I 5-6: Sep 19-Sep 29</b>	9/20, 9/27	Constituency Grammars, Parsing, Ambiguity	SLP 12.1-3, 12.5, 13.1-2
		Dependency Parsing	SLP, 14.1, 14.2, 14.3; Universal Dependencies
<b>Linguistic Structure II 7-8: Oct 3-Oct 14</b>	10/4, 10/11	Sequence Labeling for POS and NE	SLP, 8
		Labeled Corpus Example 2	SLP, 8.2, 12.4; LDC Penn Treebank
		Information Extraction, Chunking, and Additional Text Corpora (Penn Treebank, CONLL)	Natural Language Toolkit (NLTK), 0, 2, 7
		HMMs, Viterbi, Conditional Random Fields	SLP, 8.4, 8.5
		Machine Learning applied to NLP in Python	Scikit Learn, CRFSuite
<b>Semantics I 9-10: Oct 17-Oct 27</b>	10/18, 10/25	Vector Semantics and Embeddings	SLP, 6.1, 6.2, 6.3
		Cosine, Word2vec, TFIDF Similarity	SLP, 6.4, 6.5, 6.8, Slor, ElasticSearch

		Jaccard Similarity	A Simple Explanation of the Jaccard Similarity Index
		Precision, Recall, F-Measure	SLP, 4.7
<b>Semantics II 11-13: Oct 31-Nov 18</b>	11/1, 11/8, 11/15	Semantic Role Labeling and multilingual projection	SLP 19
		Contextualized Embeddings	SLP, 11.1, 11.2
<b>Applications, Eval 14-15: Nov 21-Dec 2</b>			
	[11/22], 11/29	Machine Translation	10.1 and additional notes
		Metrics: Bleu, ROUGE, etc.	SLP, 10.8.1, 10.8.2, and additional notes
		Social Computing (e.g., chatbots/conversational agents, sentiment/stance detection)	TBD
<b>Review/Final Meeting</b>			
	12/6		
<b>16: Final Exam Week</b>			
	TBD		

### **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 brief (5 to 15 min) video snippets, approximately 5-6 per module. 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) relevant to the next class period by 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 projects and homeworks; (c) discussion of the upcoming module by way of additional “thought questions.” In order 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.

**Covid-19 Response:** The Covid-19 guidelines below are adopted to maintain your learning environment and to enhance the safety of our in-classroom [F2F] interactions:

- UF is recommending that members of our campus community and visitors wear masks in indoor public settings and on public transportation. <https://coronavirus.ufl.edu/university-updates/june-15-the-campus-brief.html>
- Reusable, disposable, N95 and KN95 masks are available for purchase at the UF Bookstore. Disposable and reusable masks are available for purchase at all Gator Dining locations. Additionally, masks are available for purchase from 50 vending machines across campus.
- If you are experiencing COVID-19 symptoms, click here for guidance from the CDC on symptoms of coronavirus: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>

- Click here for UF Health guidance on what to do if you have been exposed to or are experiencing Covid-19 symptoms: <https://coronavirus.ufhealth.org/screen-test-protect-2/frequently-asked-questions/covid-19-exposure-and-symptoms-who-do-i-call-if/>

**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 (attendance is not mandatory).
- The 3-hour class time will be recorded (see the disclaimers and notice that follow).

Disclaimer #1: Many Zoom sessions are being held campus-wide to deliver courses and the demand for recorded materials is high. It is possible there will be delays while completing the recording process. *I will post live classroom recordings as soon possible [under the corresponding Canvas Module],* however the way to ensure that you see the classroom activities is to participate in the synchronous FTF classroom meeting on Tuesdays.

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 textual "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. Please note, 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 an assessment. The Dean of Students Office ([www.dso.ufl.edu](http://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"**. If the letter does not state this, then 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:** Makeups or extensions for graded class activities are provided given appropriate documentation is presented in a timely fashion; however, keep in mind that flexibility is already built into the flipped classroom schedule, to avoid the need for makeups or extensions. For example, most assessments (including exams) are applied in a time period that you select within a multi-day window. This approach allows you to plan ahead and work on assessments within a timeframe that easily accommodates scheduling constraints.

**Late Assignments** (this does not apply to Exams, which must be completed by the assigned date/time unless the absence is excused): No work can be accepted after the solution is discussed during the relevant class period. If the solution has not been discussed and you submit one day late, there will be a penalty of up to 50% on the graded work. No work that is two days late can be accepted. Note, if the assignment is due on a Fri at noon, then at 12:01PM that same day (up until 11:59AM Sat) the assignment is one day late.

**Questions outside of class time:** It is possible to set up a zoom meeting (via Canvas) with the professor or TA for private discussions that cannot be handled during a public classroom session. In such cases, it is expected that privacy is needed, e.g., to discuss a concern regarding a grade. Detailed technical questions are best addressed by leveraging the flipped classroom format. 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 is intended to serve as a much 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 also reduces redundancy, as many students have the same questions of interest to the entire class. 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 homeworks, 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 Assignments, and to review the highlights of graded assignments and exams.

**Exams:** There will be three exams. Following UF standards during the COVID-19 F2F and online course protocol, exams will be administered using Canvas *Quizzes* combined with a proctoring environment such as Honorlock. Exams 1 and 2 (each 12% of your final grade) and Exam 3 (15% of your final grade) are open book, with access to any resource made available during the course, and will be held during a 90 min time period that you select within a two-day window.

**Assignments:** There are two types of assignments: (a) Homeworks (HW) and (b) Projects. Both are electronically administered and include structured questions that are automatically graded, alongside unstructured questions that are human inspected and graded. Access to resources specified in assignments is strongly encouraged. The details of assignment submissions will be given within each assignment specification.

(a) Homeworks address foundational elements of natural language processing and employ the Canvas *Quiz* format. **One attempt** is allowed for each HW. (Download PDF version, and do HW outside of Canvas, then resume/submit inside Canvas **once** by deadline). With some exceptions due to holidays and other timing constraints, HWs are generally assigned on Mon or Tues (12:01am) and are due on Fri (noon) of the following week, with multiple attempts allowed (until the deadline). One exception is HW1, which is due Sep 2 at noon but is made available in Canvas on Wed Aug 24 (12:01am), i.e., the

Activity	Total Points	Final Grade Percentage
<i>Total Quizzes (with videos)</i>	10	5%
<i>Assignments:</i>		
• <i>Homework 1</i>	10	5%
• <i>Homework 2</i>	15	7.5%
• <i>Homework 3</i>	15	7.5%
• <i>Project 0</i>	12	6%
• <i>Project 1</i>	20	10%
• <i>Project 2</i>	20	10%
• <i>Project 3</i>	20	10%
<i>Exam 1</i>	24	12%
<i>Exam 2</i>	24	12%
<i>Exam 3</i>	30	15%
<b>Total</b>	<b>200</b>	<b>100%</b>

first day of classes for the fall semester (prior to our first in-class session). HW1 is worth 5% and HW 2 and 3 are each worth 7.5% of your final grade.

(b) Projects involve programming in Python using a Jupyter notebook format (e.g., Google Colab), coupled with both structured and unstructured responses to questions about central aspects of the programming assignment. Projects will employ a combination of both automatic grading and human inspection and grading of unstructured responses. Projects are generally assigned on Mon or Tues (12:01am) and are due on Thu or Fri (noon) of the following week. The first project (Project 0) is a warm-up exercise worth 6% of your final grade. The other three projects (Projects 1, 2, 3) are each worth 10% of your final grade.

**Videos and Quizzes:** The “flipped classroom” design is conducted via modules, each with a set of video snippets, coupled with quizzes containing practice exercises in the Canvas *Quiz* format. The sum of all quizzes counts for 5% of your final grade (each week’s quizzes contribute approximately 0.4%). These exercises are intended to highlight key elements of the videos and to help you in your studies and preparation for upcoming assessments. There are an **unlimited number of attempts** for each quiz (until the deadline). It is *highly* recommended that you watch the videos carefully, download and review the associated lecture notes (and/or the transcript for each video as needed), and then take the associated quizzes, repeating as many times as needed to get a perfect score for each one. It is expected that this will be done **before** you attend the class period where these are discussed (i.e., by 3pm every Tues), starting with those released the first day of the semester (August 24, 12:01am). The class period is a perfect opportunity for an interactive session that covers any questions or issues that emerged while you were watching the videos and taking the quizzes.

### **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/>

<b>Grade Points</b>	<b>Letter Grade</b>	<b>Highest</b>	<b>Lowest</b>
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

### **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.

### ***Students Requiring Accommodations***

Students with disabilities who experience learning barriers and would like to request accommodations should connect with the Disability Resource Center <https://disability.ufl.edu/students/get-started/> (352.392.8565, <https://disability.ufl.edu/>). Once registered, the DRC provides documentation to instructors. Students should contact instructors to ensure documentation has been received as well as to discuss accommodation specifics so that everything can be set up. Students with disabilities should follow this procedure as early as possible in the semester, delaying may affect the ability to schedule the accommodation. All exams, quizzes, and so on administered through the DRC office must follow the course exam structure, being completed on the same date and time as the course exam.

### ***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/>.

### ***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/policies/student-honor-code-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 broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

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 Graduate Program Coordinator
- Robin Bielling, Director of Human Resources, 352-392-0903, [rbielling@eng.ufl.edu](mailto:rbielling@eng.ufl.edu)
- Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, [taylor@eng.ufl.edu](mailto:taylor@eng.ufl.edu)
- Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, [nishida@eng.ufl.edu](mailto: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](mailto: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:** <http://www.counseling.ufl.edu/cwc>, 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](mailto: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](mailto:Learning-support@ufl.edu). <https://lss.at.ufl.edu/help.shtml>.

**Career Resource Center**, Reitz Union, 392-1601. Career assistance and counseling. <https://www.crc.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://www.dso.ufl.edu/documents/UF\\_Complaints\\_policy.pdf](https://www.dso.ufl.edu/documents/UF_Complaints_policy.pdf).

**On-Line Students Complaints:** <http://www.distance.ufl.edu/student-complaint-process>.