

Natural Language Processing with Python

CIS 4930

Section: 113A [F2F], 113B [Online 100%]

Class Numbers: 19999 [F2F], 26884[Online 100%]

Class Periods: MWF 3rd period [09:35 to 10:25 AM EST]

Location: MCCA 1142 | Zoom [see daily Modules in Canvas]

Academic Term: Spring 2021

Instructor:

Pete Dobbins

CSE E474

pjd@cise.ufl.edu

352.294.6685

Virtual Office Hours [VOH]: MW 4th period [10:40 – 11:30 AM], R 10th period [05:10 – 06:00 PM]

- Per UF policy, all office hours will continue to be held virtually via Zoom.
- Here is my Zoom link: <https://ufl.zoom.us/j/4755418272>
- VOH will begin Monday, January 18, 2021.
- During the **first week of classes, Wednesday, January 13, 2020 4th period** [10:40 – 11:30 AM EST] I will hold VOH to get things kicked off. If you any questions, please swing by the Zoom session and we will chat.

Teaching Assistant/Peer Mentor/Supervised Teaching Student:

Please contact TAs/PMs through the course MS Teams chat. Announcements will be posted in MS Teams. Invitations to our Team will be sent out after drop/add has ended. Within our Canvas course shell you will find the VOH schedule (Modules/General Course Info/Weekly VOH Schedule) and Zoom links for each staff member.

- Michael Kummer
- George Pu

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 practical application and theory of Natural Language Processing. Practically, we are motivated to study natural language to better understand how to search abstract data sets (for example the *deep-web*) for query answer retrieval. Among the theoretical topics included are: regular expressions, collocations, n-grams, syntactic parsing, word sense disambiguation, lexical acquisition, Markov models, context-free grammars, and information retrieval. **Python** will be introduced and used to illustrate course concepts. Project solutions will be developed in **Python**. Weekly course topics are provided on the calendar pages of the course website.

Students will be able to apply their knowledge to:

- evaluate the performance of natural language applications
- search abstract data sets (for example the *deep-web*) for query answer retrieval
- develop, implement, and evaluate their own programmatic solutions to natural language problems
- apply standard mathematical and statistical principles to natural language problems

Materials and Supply Fees

There are no materials or supply fees for this course beyond the textbook.

Required Textbooks and Software

- The second edition of *Natural Language Processing with Python* is published here: <http://www.nltk.org/book/>.

Supplemental Textbooks

- *Hands-on Machine Learning with Scikit-Learn & TensorFlow*, Second Edition, Aurelien Geron, O'Reilly Media, 2019, ISBN 1-492-03264-6
- *Learning Python*, Fourth Edition, Mark Lutz, 2009, O'Reilly Media, ISBN 0-596-15806-8.
- *Foundations of Statistical Natural Language Processing*, First Edition, Manning & Schuetze, 1999, ISBN 0-262-13360-1.
- *Speech and Language Processing*, Second Edition, Jurafsky & Martin, 2009, Prentice Hall, ISBN 0-131-87321-0.

Resources

Recommendations for other resources are given here.

- *Natural Language Processing with Python*, First Edition, Bird, Klein, & Loper, 2009, O'Reilly Media, ISBN 0-596-51649-5. (this is the first edition of the textbook; it is not updated for NLTK 3 and Python 3)
- *Errata* from the June 2009 printing: <http://nltk.googlecode.com/svn/trunk/nltk/doc/book/errata.txt>.
- Website: <http://www.nltk.org/>.
- Python: <http://www.python.org/>.
- Learn and Practice Python: <http://inventwithpython.com/>.
- NLP on Wikipedia: 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. Note, under Modules, individual lectures are listed using the format: **year / month / day | L##: General Topic**.

Dates	Week	Topics	Reading
2021 / 01 / 10 - 01 / 16	1	Syllabus, Introduction, Installation	Syllabus 1.1 - 1.2
2021 / 01 / 17 - 01 / 23	2	Bigrams, Collocations, NLTK Functions, Frequency Distribution, Testing vs Training, Lists, Overfitting, Corpora	1.3 - 1.8 2.1
2021 / 01 / 24 - 01 / 30	3	Conditional Frequency Distribution, Classifying Antagonizing Text, Wordnet, Similarity Measures, Grammatical Definitions	2.1 - 2.8
2021 / 01 / 31 - 02 / 06	4	Zipf's Law, Exam 1, File IO	
2021 / 02 / 07 - 02 / 13	5	File IO, Regular Expressions	3.1 - 3.12
2021 / 02 / 14 - 02 / 20	6	Problem Solving in Python, Sample Exercises	4.1 - 4.11
2021 / 02 / 21 - 02 / 27	7	Finite State Automata, Collocations, Parts of Speech, Variance, t Test	5.1 - 5.5
2021 / 02 / 28 - 03 / 06	8	Review, Exam 2, Mean Differential, Chi-Square Test	5.5 - 5.10
2021 / 03 / 07 - 03 / 13	9	Mean Differential, Chi-Square Test, Parts of Speech	
2021 / 03 / 14 - 03 / 20	10	Confusion Matrix, Precision, Recall, Cross Validation, Supervised Classification	6.1 - 6.10
2021 / 03 / 21 - 03 / 27	11	Bayesian Analysis, Feature Extraction, Classifiers: Decision Tree, Max Entropy	7.1 - 7.9

2021 / 03 / 28 - 04 / 03	12	Exam 3	
2021 / 04 / 04 - 04 / 10	13	Chunking	
2021 / 04 / 11 - 04 / 17	14	Context Free Grammars	8.1 - 8.9
2021 / 04 / 18 - 04 / 24	15	Exam 4, Classes End	-
2021 / 04 / 25 - 05 / 01	16	Final Exams Week	-

Attendance Policy, Class Expectations, and Makeup Policy

General policies can be found here: <http://www.cise.ufl.edu/~pjd/admin/policies/course.html>. You are expected **know and follow** all of these policies. In order to be successful in this course, you should attend lecture. There is no replacement to the experience of seeing the material presented during lecture.

Live lectures will be *presented synchronously* in our assigned classroom [F2F] and simultaneously streamed virtually [online] via **Zoom: MWF 3rd period [09:35 to 10:25 AM EST]**. University policy states that you should attend using the medium [F2F or online] of your registered section. If you would like to change sections, please adjust your schedule accordingly through course registration [<https://one.ufl.edu/>] or speak with CISE Academic Advising [<https://www.cise.ufl.edu/academics/advising-information/>].

In response to COVID-19, the following policies and requirements are in place to maintain your learning environment and to enhance the safety of our in-classroom [F2F] interactions.

- You are required to wear approved face coverings at all times during class and within buildings. Following and enforcing these policies and requirements are all of our responsibility. Failure to do so will lead to a report to the Office of Student Conduct and Conflict Resolution.
- This course has been assigned a physical classroom with enough capacity to maintain physical distancing (6 feet between individuals) requirements. Please utilize designated seats and maintain appropriate spacing between students. Please do not move desks or stations.
- Sanitizing supplies are available in the classroom if you wish to wipe down your desks prior to sitting down and at the end of the class.
- Follow your instructor's guidance on how to enter and exit the classroom. Practice physical distancing to the extent possible when entering and exiting the classroom.
- If you are experiencing COVID-19 symptoms ([Click here for guidance from the CDC on symptoms of coronavirus](#)), please use the UF Health screening system and follow the instructions on whether you are able to attend class. [Click here for UF Health guidance on what to do if you have been exposed to or are experiencing Covid-19 symptoms.](#)

Here are points of significance for our lecture process this semester.

- You are responsible for all material covered during lecture.
- Lecture attendance will not be taken (attendance is not mandatory).
- Lectures 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 lecture recordings as soon possible [under the corresponding Canvas Module]*, however if you want to ensure that you see the lecture as soon you are able, you should participate in the synchronous live lecture.

Disclaimer #2: the Zoom platform records these sessions and I cannot guarantee they will not experience errors in the recording process. I have not observed this to be a significant problem. However, you should always be aware that attending the live lecture session is always the best way to ensure you are able to view the complete lecture content.

Notice: Our lectures will be 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 you “chat” to be recorded, you must send it through the private channel directly and only to the instructor. Please note, since lecture is a public setting, private chats may or may not always be answerable during the live lecture. VOH are provided throughout the week to accommodate any private discussions that cannot be handled during a lecture session. As in all courses, unauthorized recording and unauthorized sharing of recorded materials is prohibited.

Makeups or extensions for graded class activities are provided given appropriate documentation is presented in a timely fashion. Excused absences must be consistent with University policies in the undergraduate catalog (<https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/>) and meet University requirements regarding excused absences. There are very few events which would impede your participation in a graded activity warranting an exception, that you would not be aware of well before [at least a few hours] the start of the activity. The Dean of Students Office (www.dso.ufl.edu) provides rules, guidance, and approval for excuse documentation. Once an excuse has been approved by DSO, appropriate documentation will be provided to your instructors. *All excused exams will be made up at the same time, on the same cumulative exam, for a 120 minute [2-hour] timeframe, and held on Monday, April 26, 2021 between 10:00 AM EST and 12:00 PM EST.* Note, this is our UF assigned final exam period block. Any conflicts with the structured makeup follow the standard exam rules, UF approved excuse documentation must be presented in a timely fashion.

Note, Exams 1, 2, 3, and 4 will be administered during our UF assigned lecture period on the dates provided in the table below. This is structured to provide our course with the simplest format for avoiding exam conflicts. Reference the Exams section below for more details.

Late work [such as HW, this is not applicable to in other course activities or Exams which must be completed by the assigned date/time of the activity unless the absence is excused]: No work can be accepted after the solution is posted. If the solution has not been posted 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 at 11:59 PM, then at 12:00 AM the work is one day late.

Evaluation of Grades

Activity	Final Grade Percentage	Comment
<i>Exam 1</i>	10%	
<i>Exam 2</i>	12.5%	
<i>Exam 3</i>	12.5%	
<i>Exam 4</i>	15%	
<i>Homework</i>	30%	1 drop [lowest score]
<i>Project</i>	20%	
Total	100%	

Exams

There will be four exams, totaling 50% of your final grade. **Exam 1** will be **10%**, **Exam 2** will be **12.5%**, **Exam 3** will be **12.5%**, and **Exam 4** will be **15%** of your final grade. 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 will be held during our UF assigned Monday lecture period. You will have *60 minutes* [1-hour] to complete the exam, **starting at 09:25 AM EST** and **ending at 10:35 AM EST** on the dates given in this table.

Exam	Grade %	Day	Date
1	10%	Wednesday	2021 / 02 / 03
2	12.5%	Wednesday	2021 / 03 / 03
3	12.5%	Wednesday	2021 / 03 / 31
4	15%	Wednesday	2021 / 04 / 21

Homework [HW]

Homework will be completed *individually*, including written problems and the implementation of Python programs. Your *lowest* homework score will automatically be *dropped*. The details of homework submission will be given within each assignment specification (for example through a soft copy via [e-Learning \[Canvas\]](#) by 11:59 PM on the due date assigned within the project specification).

Project

Using the concepts we discuss, you will create a solution to a natural language problem you propose. The project is split into 3 parts: *Proposal*, *Evaluation Plan*, and *Results*. Proposals must be approved by course staff before submission, allowing us to help you identify an appropriate project.

Completing the project does not require you work in a group. Optionally, you are allowed to work in a group of size 2. Groups are self-managed [you pick your partner]. It is each student's responsibility to manage their individual effort. If you find your group is not working effectively, you are allowed a one-time shift to working individually. This request must be made by March 31, 2021 at 5 PM

Part	Activity	Grade %	Day	Date
1	Proposal	5%	Tuesday	2021 / 03 / 16
2	Evaluation Plan	5%	Tuesday	2021 / 04 / 13
3	Results	10%	Tuesday	2021 / 04 / 27

Practice Exercises

There will be practice exercises assigned during lecture that are not associated with a graded homework. These exercises are supplemental to the course lectures and graded materials, helping you in your studies and preparation. These exercises will not be collected, however if you would like to discuss your solutions with us, please do so. It is *highly* recommended that you complete and participate in non-graded course activities: *lecture* and *practice exercises*, in addition to the graded activities: *exams* and *homework*.

Grading Policy

The range used to calculate your final letter grade in our course will be no harsher than this grading scale provided in the following 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.

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/UGRD/academic-regulations/grades-grading-policies/>

Re-Grade Requests

You must return exams/projects/quizzes to the person grading the activity in question, along with a **printed (not handwritten) attachment** stating the reasons for the re-grade request. Do not write anything on the exam/project/quiz itself or detach (un-staple) the pages from each other. If you do so, your exam/project/quiz will not be re-graded. Note, when requesting a re-grade, the whole exam/project/quiz is subject to the re-grade, not just the portion you specify.

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 setup. 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
- Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.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: <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

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://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.

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