Cover Sheet: Request 15990

LIN 4XXXC – Introduction to Computational Linguistics

Info

Process	Course|New|Ugrad/Pro
Status	Pending at PV - University Curriculum Committee (UCC)
Submitter	Kevin Tang tang.kevin@ufl.edu
Created	3/18/2021 3:25:01 PM
Updated	4/7/2021 8:55:51 PM
Description of request	The field of linguistics is becoming more computational and quantitative. Especially with current easy access to big data, and growing interest in data mining, machine learning and human-computer interfaces, there is a growing need for linguists to have computational skills. This course distinguishes itself from courses offered in other departments in terms of its application to problems and data sets in the language sciences. This course is offered as an advanced undergraduate course (4000-level), since knowledge of linguistic terminology and issues is assumed.

Actions

Step	Status	Group	User	Comment	Updated
Department	Approved	CLAS - Linguistics 16300000	Eric Potsdam		3/28/2021
External-Consult_signed_Gilbert.pdf					3/18/2021
College	Conditionall Approved	CLAS - College of Liberal Arts and Sciences	Joseph Spillane	• Course Objectives need to be revised. Need to be measurable. Suggestion: “Describe” rather than “Be familiar with”. • Transcript Title error in 4XXX version: Intro to Computational Ling • Late assignments will need to be softened to “excused circumstances” from “extreme”.	4/7/2021
No document changes					
Department	Approved	CLAS - Linguistics 16300000	Eric Potsdam	Changes requested by College approver (course objectives, late work, transcript title) were made	4/7/2021
LIN4XXX Intro to Comp Lx.pdf LIN5XXX Intro to Comp Lx.pdf					4/7/2021 4/7/2021
College	Approved	CLAS - College of Liberal Arts and Sciences	Joseph Spillane		4/7/2021
No document changes					
University Curriculum Committee	Pending	PV - University Curriculum Committee (UCC)			4/7/2021
No document changes					
Statewide Course Numbering System					
No document changes					
Office of the Registrar					

Step	Status	Group	User	Comment	Updated
No document changes					
Student Academic Support System					
No document changes					
Catalog					
No document changes					
College Notified					
No document changes					

Course|New for request 15990

Info

Request: LIN 4XXXC – Introduction to Computational Linguistics
Description of request: The field of linguistics is becoming more computational and quantitative. Especially with current easy access to big data, and growing interest in data mining, machine learning and human-computer interfaces, there is a growing need for linguists to have computational skills. This course distinguishes itself from courses offered in other departments in terms of its application to problems and data sets in the language sciences. This course is offered as an advanced undergraduate course (4000-level), since knowledge of linguistic terminology and issues is assumed.
Submitter: Eric Potsdam potsdam@ufl.edu
Created: 4/7/2021 8:43:01 PM
Form version: 5

Responses

Recommended Prefix
Enter the three letter code indicating placement of course within the discipline (e.g., POS, ATR, ENC). Note that for new course proposals, the State Common Numbering System (SCNS) may assign a different prefix.

Response:
LIN

Course Level
Select the one digit code preceding the course number that indicates the course level at which the course is taught (e.g., 1=freshman, 2=sophomore, etc.).

Response:
4

Course Number
Enter the three digit code indicating the specific content of the course based on the SCNS taxonomy and course equivalency profiles. For new course requests, this may be XXX until SCNS assigns an appropriate number.

Response:
XXX

Category of Instruction
Indicate whether the course is introductory, intermediate or advanced. Introductory courses are those that require no prerequisites and are general in nature. Intermediate courses require some prior preparation in a related area. Advanced courses require specific competencies or knowledge relevant to the topic prior to enrollment.

Response:
Advanced

- *1000 level = Introductory undergraduate*
- *2000 level = Introductory undergraduate*
- *3000 level = Intermediate undergraduate*
- *4000 level = Advanced undergraduate*
- *5000 level = Introductory graduate*
- *6000 level = Intermediate graduate*
- *7000 level = Advanced graduate*
- *4000/5000= Joint undergraduate/graduate*

• 4000/6000= Joint undergraduate/graduate

**Joint undergraduate/graduate courses must be approved by the UCC and the Graduate Council)*

Lab Code
Enter the lab code to indicate whether the course is lecture only (None), lab only (L), or a combined lecture and lab (C).

Response:
C

Course Title
*Enter the title of the course as it should appear in the Academic Catalog. There is a 100 character limit for course titles. *

Response:
Introduction to Computational Linguistics

Transcript Title
Enter the title that will appear in the transcript and the schedule of courses. Note that this must be limited to 30 characters (including spaces and punctuation).

Response:
Intro Computational Ling

Degree Type
Select the type of degree program for which this course is intended.

Response:
Baccalaureate

Delivery Method(s)
Indicate all platforms through which the course is currently planned to be delivered.

Response:
On-Campus

Co-Listing
Will this course be jointly taught to undergraduate, graduate, and/or professional students?

Response:
Yes

Co-Listing Explanation
Please detail how coursework differs for undergraduate, graduate, and/or professional students. Additionally, please upload a copy of both the undergraduate and graduate syllabus to the request in .pdf format. For more information please see the Co-Listed Graduate Undergraduate Courses Policy.

Response:
Undergraduate: Homework: 30%, Final project: 35%
Graduate: Homework 20%, Final project 45% for which the students would be asked to select a computational linguistics conference venue as a target outlet and prepare as a group a conference paper (typically 7 pages, two columns) as well as an abstract for submission (this requirement doesn't exist for the Undergraduate level).

Effective Term
Select the requested term that the course will first be offered. Selecting "Earliest" will allow the course to be active in the earliest term after SCNS approval. If a specific term and year are selected, this should reflect the department's best projection. Courses cannot be implemented retroactively, and therefore the actual effective term cannot be prior to SCNS approval, which must be obtained prior to the first day of classes for the effective term. SCNS approval typically requires 2 to 6 weeks after approval of the course at UF.

Response:
Earliest Available

Effective Year
Select the requested year that the course will first be offered. See preceding item for further information.

Response:
Earliest Available

Rotating Topic?
Select "Yes" if the course can have rotating (varying) topics. These course titles can vary by topic in the Schedule of Courses.

Response:
No

Repeatable Credit?
Select "Yes" if the course may be repeated for credit. If the course will also have rotating topics, be sure to indicate this in the question above.

Response:
No

Amount of Credit
Select the number of credits awarded to the student upon successful completion, or select "Variable" if the course will be offered with variable credit and then indicate the minimum and maximum credits per section. Note that credit hours are regulated by Rule 6A-10.033, FAC. If you select "Variable" for the amount of credit, additional fields will appear in which to indicate the minimum and maximum number of total credits.

Response:
3

S/U Only?
Select "Yes" if all students should be graded as S/U in the course. Note that each course must be entered into the UF curriculum inventory as either letter-graded or S/U. A course may not have both options. However, letter-graded courses allow students to take the course S/U with instructor permission.

Response:
No

Contact Type
Select the best option to describe course contact type. This selection determines whether base hours or headcount hours will be used to determine the total contact hours per credit hour. Note that the headcount hour options are for courses that involve contact between the student and the professor on an individual basis.

Response:
Regularly Scheduled

- *Regularly Scheduled [base hr]*
- *Thesis/Dissertation Supervision [1.0 headcount hr]*
- *Directed Individual Studies [0.5 headcount hr]*
- *Supervision of Student Interns [0.8 headcount hr]*
- *Supervision of Teaching/Research [0.5 headcount hr]*
- *Supervision of Cooperative Education [0.8 headcount hr]*

Contact the Office of Institutional Planning and Research (352-392-0456) with questions regarding contact type.

Weekly Contact Hours
Indicate the number of hours instructors will have contact with students each week on average throughout the duration of the course.

Response:
3

Course Description
Provide a brief narrative description of the course content. This description will be published in the Academic Catalog and is limited to 500 characters or less. See course description guidelines.

Response:
Introduces the study of natural language from a computational perspective. Discusses both applied (engineering) and theoretical (cognitive) issues, ranging from speech and language technology to formal aspects of theoretical linguistic models. Covers a brief introduction to programming in Python, and the basics of Natural Language Processing and their applications.

Prerequisites
Indicate all requirements that must be satisfied prior to enrollment in the course. Prerequisites will be automatically checked for each student attempting to register for the course. The prerequisite will be published in the Academic Catalog and must be formulated so that it can be enforced in the registration system. Please note that upper division courses (i.e., intermediate or advanced level of instruction) must have proper prerequisites to target the appropriate audience for the course.
Courses level 3000 and above must have a prerequisite.
Please verify that any prerequisite courses listed are active courses.

Response:

LIN3010

Completing Prerequisites on UCC forms:

• Use “&” and “or” to conjoin multiple requirements; do not used commas, semicolons, etc.
• Use parentheses to specify groupings in multiple requirements.
• Specifying a course prerequisite (without specifying a grade) assumes the required passing grade is D-. In order to specify a different grade, include the grade in parentheses immediately after the course number. For example, "MAC 2311(B)" indicates that students are required to obtain a grade of B in Calculus I. MAC2311 by itself would only require a grade of D-.
• Specify all majors or minors included (if all majors in a college are acceptable the college code is sufficient).
• “Permission of department” is always an option so it should not be included in any prerequisite or co-requisite.
• If the course prerequisite should list a specific major and/or minor, please provide the plan code for that major/minor (e.g., undergraduate Chemistry major = CHY_BS, undergraduate Disabilities in Society minor = DIS_UMN)

Example: A grade of C in HSC 3502, passing grades in HSC 3057 or HSC 4558, and undergraduate PBH student should be written as follows: HSC 3502(C) & (HSC 3057 or HSC 4558) & UGPBH
* *

Co-requisites
Indicate all requirements that must be taken concurrently with the course. Co-requisites are not checked by the registration system. If there are none please enter N/A.

Response:
N/A

Rationale and Placement in Curriculum
Explain the rationale for offering the course and its place in the curriculum.

Response:
The field of linguistics is becoming more computational and quantitative. Especially with current easy access to big data, and growing interest in data mining, machine learning and human-computer interfaces, there is a growing need for linguists to have computational skills. This course distinguishes itself from courses offered in other departments in terms of its application to problems and data sets in the language sciences. This course is offered as an advanced undergraduate course (4000-level), since knowledge of linguistic terminology and issues is assumed.

Course Objectives
Describe the core knowledge and skills that student should derive from the course. The objectives should be both observable and measurable.

Response:
On completion of this course, students should:
• Be able to use computational linguistic tools and resources, and be able to describe how they are applied in research in both computational linguistics and other subfields
• Have a rough sense of the state of the art in this subfield
• Be able to conceptualize problems from the perspective of computational linguistics
• Be able to code in Python and use JyputerNotebook.

Course Textbook(s) and/or Other Assigned Reading
*Enter the title, author(s) and publication date of textbooks and/or readings that will be assigned. Please provide specific examples to evaluate the course and identify required textbooks. *

Response:
Daniel Jurafsky and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. isbn: 0131873210
Readings from the textbooks will be supplemented by other readings and materials throughout the semester (made available on the website for the course).

Weekly Schedule of Topics
Provide a projected weekly schedule of topics. This should have sufficient detail to evaluate how the course would meet current curricular needs and the extent to which it overlaps with existing courses at UF.

Response:
Course outline (tentative, and subject to revision)
Readings should be completed before the class date listed.
Acronyms:
• J&M: Jurafsky and Martin (2008)
• L&C: Language and Computer (2013)
• H: Hetland (2008)
• BKL: Bird et al (2009)

Week
Topic
Reading
Assignment

W1
Introduction and Syllabus, Python: Intro, LinkedIn Learning
J&M: Ch 1 (L&C Ch 1: Optional)

W2
Python: Intro, LinkedIn Learning

W3
Python: Data types & Files,Python: Control & Reg Exp, Text Normalization
H: Ch 2-4, (Ch 11: Optional); J&M: Ch 8
Programming assessment

W4
Lab: Text Processing, Edit Distance, The Noisy Channel
BKL: Ch 1,2,3; J&M: Ch 3; J&M: Ch 5, L&C: Ch2
Programming assessment

W5
The Noisy Channel; N-Grams
J&M: Ch 5, L&C: Ch 2; J&M: Ch 4
HW 1 set, programming assessment

W6
N-Grams; Lab: N-Grams
J&M: Ch 4
HW 1 due

W7
Machine Learning: Overview; Evaluation and Error analysis
T. Mitchel. (2017); Resnik & Lin (2010), Kummerfeld et al. (2012)
Project milestone 1 due

W8

Evaluation and Error analysis; Regression and Maximum Entropy
Resnik & Lin (2010), Kummerfeld et al. (2012); J&M: Ch 6
HW 2 set

W9
Regression and Maximum Entropy; Lab: Regression
J&M: Ch 6
HW 2 due, Project milestone 2 due

W10
Part of Speech tagging
J&M: Ch 5

W11
Lab: Tagging; Vector Semantics
BKL: Ch 5; TBD
HW 3 Set, Project milestone 3 due

W12
Vector Semantics; Lab: Semantics; Naive Bayes Classification
TBD
HW 3 due

W13
Sentiment Analysis; Sentiment, Affect, and Connotation
TBD
HW 4 Set

W14
Lab: Sentiment Analysis; Modelling Grammar; Modelling Intuitions
TBD
HW 4 due

W15
Ethics; Poster session
TBD
Project milestone 4 due

Full references:
Bird, Steven, Ewan Klein, and Edward Loper (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.
Hetland, Magnus Lie (2008). Beginning Python from Novice to Professional. Wiley.
Jurafsky, Daniel and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. Isbn: 0131873210.
Mitchell, T.M.: Key Ideas in Machine Learning.
http://www.cs.cmu.edu/%7Etom/mlbook/keyIdeas.pdf (2017). Accessed 28 Jan 2019
Resnik, P. & J. Lin (2010), Evaluation of nlp systems, in A. Clark, C. Fox, & S. Lappin (eds.), Handbook of Computational Linguistics and Natural Language Processing, Wiley-Blackwell, Oxford.
Kummerfeld, J. K., Hall, D., Curran, J. R., & Klein, D. (2012, July). Parser showdown at the wall street corral: An empirical investigation of error types in parser output. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 1048-1059).

Grading Scheme
List the types of assessments, assignments and other activities that will be used to determine the course grade, and the percentage contribution from each. This list should have sufficient detail to evaluate the course rigor and grade integrity. Include details about the grading rubric and percentage breakdowns for determining grades. If

participation and/or attendance are part of the students grade, please provide a rubric or details regarding how those items will be assessed.

Response:
Grading (tentative, and subject to revision)
• Class participation: 10%
Active in-class participation is a requirement of the course. 'Active participation' means that you should regularly ask thoughtful questions in class during lectures and tutorials, and participate with the group exercises during tutorials. If you are habitually absent from class, leave early without letting us know ahead of time, or are otherwise disengaged (e.g. on your smartphone), that will negatively affect your participation grade.
If you miss a class, it is up to you to borrow notes from someone, ask other students about changes to the reading/homework schedule, etc. Please don't ask me to go over what we did in class.
• Quizzes: 10%
Quizzes will be available via Canvas. They are designed to test your on-going knowledge of the course content, specifically the conceptual content.
• Programming assessment (approx. 3 pieces): 5%
• Homework (approx. 3 pieces): 30%
Assignments will be uploaded to the course website.
Assignments must be submitted via .
Late assignments will NOT be accepted, except under extreme circumstances.
Emailed assignments will NOT be accepted, except under extreme circumstances.
In general, I will distribute the assignments one week before they are due, in class and/or on e-Learning.
I will use automatic checks for overlap between your code and other students' code.
Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade.
Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
Clarification: to do well in the HWs, you should examine the HW immediately, pay attention to the number of points assigned to each section, allocate a sufficient amount of time per section, and ask for clarification of anything immediately via a Canvas Discussion thread. Typically, most of the points are allocated in later sections of the HWs, therefore you should persevere and do not give up.
• Lab work (approx. 4 pieces) : 10%
Completion of lab exercises. You should submit the answers of the lab exercises. Generally, the deadline is a week after each of the lab sessions.
Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade.
Clarification: to do well in the lab exercises, you should work on the lab exercises as much as you can in our lab sessions, and ask for clarification of anything during class and via a Canvas Discussion thread.
• Final project: 35%
Project presentations/demos: you are also required to present your work as a group. The exact format is to be determined. It will likely be a poster session open to the public.
Submission of an individual write-up of your final project (max. 10 pages) – submitted individually.
Submission of the group project files (the data, codes and presentation files) – submitted by your team representative.
The project will have four milestones.

Project milestone 1: Create your project team (max 3 people) and propose three research proposals to be developed and submit them for review.
Project milestone 2: Meet with instructors to select one proposal to develop further. Submit a draft for a developed version of this proposal
Project milestone 3: Submit the final proposal based on feedback from the instructors
Project milestone 4: Present your research as a poster
Grading scale
A: 92-100, A-: 88-91.9,
B+ 85-87.9, B: 81-84.9, B-: 78-80.9,
C+: 75-77.9, C: 71-74.9, C-: 68-70.9,
D+: 65-67.9, D: 61-64.9, D-: 58-60.9,
E: Below 58

Instructor(s)
Enter the name of the planned instructor or instructors, or "to be determined" if instructors are not yet identified.

Response:
Kevin Tang

Attendance & Make-up
Please confirm that you have read and understand the University of Florida Attendance policy.
A required statement statement related to class attendance, make-up exams and other work will be included in the syllabus and adhered to in the course. Courses may not have any policies which conflict with the University of Florida policy. The following statement may be used directly in the syllabus.

• Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at:
https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx

Response:
Yes

Accomodations
Please confirm that you have read and understand the University of Florida Accommodations policy.
A statement related to accommodations for students with disabilities will be included in the syllabus and adhered to in the course. The following statement may be used directly in the syllabus:

• Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc/) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

Response:
Yes

UF Grading Policies for assigning Grade Points
Please confirm that you have read and understand the University of Florida Grading policies.
Information on current UF grading policies for assigning grade points is require to be included in the course syllabus. The following link may be used directly in the syllabus:

• *https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx*

Response:
Yes

Course Evaluation Policy

Course Evaluation Policy
Please confirm that you have read and understand the University of Florida Course Evaluation Policy.
A statement related to course evaluations will be included in the syllabus. The following statement may be used directly in the syllabus:

• *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/public-results/. 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/.*

* *

Response:
Yes

UF UNIVERSITY of FLORIDA

UCC: External Consultations

External Consultation Results (departments with potential overlap or interest in proposed course, if any)

Department	Name and Title
CISE	Juan Gilbert, CISE Chair
Phone Number	**E-mail**
352-392.1527	juan@ufl.edu

Comments

The CISE Department does not have any issues with these courses moving forward.

Department	Name and Title
Phone Number	**E-mail**

Comments

Department	Name and Title
Phone Number	**E-mail**

Comments

LIN4XXX Introduction to Computational Linguistics Instructor: Dr. Kevin Tang

General Information

Course description

Computational linguistics is the study of natural language from a computational perspective. It encompasses both applied (engineering) and theoretical (cognitive) issues, ranging from speech and language technology to formal aspects of theoretical linguistic models.

This course is for YOU if you have ever wondered: How your phone can do predictive texting? How does a computer understand the syntax of a sentence? How Amazon can process billions of reviews? How to automatically process large corpora (big data)? How to discover linguistic structures from language data automatically? How to model our linguistic intuitions of grammaticality?

This course is for YOU if you want to build a foundation for areas such as data science, the tech industry https://www.linguisticsociety.org/resource/webinar-linguists-and-linguistics-tech or for applying for a Master's in Computational Linguistics.

This course is for YOU if you want to create a computational linguistics project like these ones by last year's students: https://slam.lin.ufl.edu/2020/04/20/intro2compling-symposium-spring2020/.

In this class, we will survey various topics and tasks in computational linguistics. While we will cover some of the basics of Natural Language Processing (which we will consider a separate subfield), this class will not focus on one specific approach (such as deep learning). Students in this class are expected to have a background in either computer science or linguistics, but not necessarily both. Expect this class to be difficult at times and easy at others. We hope to offer something new and interesting for everyone.

Please also check out this page to learn more about computational linguistics at UF: https://slam.lin.ufl.edu/computational-linguistics/

Objectives

On completion of this course, you should:

- Be able to use computational linguistic tools and resources, and be able to describe how they are applied in research in both computational linguistics and other subfields

- Have a rough sense of the state of the art in this subfield
- Be able to conceptualize problems from the perspective of computational linguistics
- Be able to code in Python and use JyputerNotebook.

Time and place

WHERE: TBA

WHEN: TBA

Instructor information

Instructor: Dr. Kevin Tang

– https://slam.lin.ufl.edu/

Email:

- `tang.kevin@ufl.edu`

Offices:

- 4017 Turlington Hall, Gainesville, FL 32611-5454

Office hours:

- TBD or by appointment.
- For information on what are office hours and how to make use of them properly see: http://lsc.cornell.edu/wp-content/uploads/2015/10/What-Are-Office-Hours.pdf.

Requirements

Prerequisites

LIN3010 (Introduction to Linguistics).

Programming training

Programming is not the focus of this course, but knowing how to program is an essential skill needed to do computational linguistics. I have engineered a number of ways to get you all up to speed with programming.

- **The first three weeks** will cover a brief introduction to programming in Python, with a particular focus on learning how to process written text, and JyputerNotebook (an interactive computational environment).
- **Two free online courses** will be assigned as assignments in a flipped class room format.
- After these first few weeks, you should have **a basic foundation of Python** to be able to tackle non-coding related course components.
- The course will include a complementary survey of the **Python Natural Language Toolkit (NLTK)** which lowers the need of coding for certain topics of the course.

Note that the programming training will not be entirely comprehensive as one can never finish

learning to code. You need to be prepared to be **highly motivated** when it comes to learning how to code using Python. Learning to code is *not* like learning a language or learning to do math (see https://news.mit.edu/2020/brain-reading-computer-code-1215). This means you should complete all the assigned online courses and exercises, and make use of the class time, Canvas discussion forums, office hours, and any supplementary resources provided (e.g., the Python textbook).

To stay motivated, it is important to consider why should you learn to code and how coding can improve your career prospect and open doors. Here are very cool animated videos on the basic concepts of coding by Karlie Kloss and TedTalk-Ed.

- "Coding is a superpower": https://www.youtube.com/watch?v=Bwiln7v0fdc
- "Karlie Kloss Explains How Computers Work": https://www.youtube.com/watch?v=_zeV-xWSJoU
- "Variables, Functions, and Arrays with Karlie Kloss": https://www.youtube.com/watch?v=MXkGONNYKHo
- "Think like a coder" (10 episodes): https://youtu.be/KFVdHDMcepw

To understand your suitability, I encourage you to check out the two free online courses that we will assign as assignments during the first three weeks. As always, you can always **reach out to me at** `tang.kevin@ufl.edu` **to discuss your suitability**.

- Note: As a UF student, the courses below are free (see instructions here: https://elearning.ufl.edu/supported-services/linkedin-learning/)
- Course 1: Python Essential Training (estimated duration: 4h 45m) by Bill Weinman https://www.linkedin.com/learning/python-essential-training-2
- Course 2: Introducing Jupyter by Josh McQuiston (estimated duration: 53m) https://www.linkedin.com/learning/introducing-jupyter/

Course website

- There is a course website on Canvas
- Let me know if you are not on the site.

Textbook and Reading list

Main textbook (Second edition):

- Jurafsky, Daniel and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. Isbn: 0131873210.
- Make sure you get the second edition! Within the second edition, the hardcover “US edition" is preferred, but you can also buy the “international edition" (which might be cheaper). There’s also an ebook version of the second edition (which again might be cheaper). They differ mostly only in a few exercises at the end of each chapter.

Supplementary textbook (Freely available online):

- Freely available online: https://www.nltk.org/book/

- Dickinson, Brew, and Meuers (2013), *Language and Computers*
- Available for reading online through the UF library (and download up to 51 pages)

Python textbook (Available at UF):

- Available for download as an ebook through the UF library

Programming Web Resources (Available at UF):

- Note: As a UF student, the LinkedIn courses below are free (see instructions here: https://elearning.ufl.edu/supported-services/linkedin-learning/)
- Course 1: Python Essential Training (estimated duration: 4h 45m) by Bill Weinman https://www.linkedin.com/learning/python-essential-training-2
- Course 2: Introducing Jupyter by Josh McQuiston (estimated duration: 53m) https://www.linkedin.com/learning/introducing-jupyter/
- Course 3 (Optional, but recommended): https://www.linkedin.com/learning/python-functions-for-data-science/
- Additionally, there are numerous online resources for Python, and you may find them a convenient supplement to the Hetland book, especially if you're an experienced programmer. https://docs.python.org/3/tutorial/, and https://www.codecademy.com/learn/learn-python.
- If you have a specific problem or question, Stack Overflow is a good place to look for answers: stackoverflow.com. I recommend searching the site for your question before asking it yourself.
- Learning how to google specific questions to debug is highly recommended (http://letmegooglethat.com/)
- Readings from the textbooks will be supplemented by other readings and materials throughout the semester (made available on the website for the course).

Course requirements

1. **Lectures and Reading:** The content of the course will be presented through lectures and reading assignments. You should attend lectures and complete the reading assignment by the date of the lecture with which the reading is associated. The material in the lectures and readings will not be identical, so you will need to both attend lectures and do the readings to succeed. Homework and quizzes will presume familiarity with both the readings and the lectures.
2. **Quizzes:** There will be a number of short quizzes in this course which are designed to test your conceptual knowledge of computational linguistics.
3. **Programming assessments**: There will be a number of programming assignments.
4. **Lab exercises**: There will be a number of lab sessions, indicated in the schedule below, which will allow you to practice the skills you have learned in the course in a supervised environment. They will test your knowledge of the material we discuss in class – a combination of programming and conceptual knowledge. **You are encouraged to work**

in a group with these lab exercises. In my experience, students benefit a lot from working with each other.

5. **Homework assignments:** There will be a number of homework assignments (on the order of 3 or 4) in this course, which will involve programming. **You are encouraged to work in a group with these lab exercises.** These assignments will require you to *implement algorithms* from computational linguistics, t*est them on data sets*, and suggest and *explore potential improvements*.
6. **Final Project**: **You will work in a group of 2-3 people on a final project** that builds in some way on the material we cover in class and/or connects with related literature. The amount of work that I expect from you will, of course, vary by the number of people who contribute, but you may find it helpful and interesting to work with others who have backgrounds different from yours. The range of allowable topics is very flexible: it may relate to a personal passion of your (e.g., comic books) or to an area of research. I would encourage you to try to find a topic that inspires you (and your partners), since this is something that you will be spending a good deal of time on.

Once you have identified a topic with your group, you should schedule an appointment with me to finalize the idea. Your project must include a computational implementation of some sort, and must be presented in a written report that explains what you have done, how it relates to past work, and what you have learned from your results. You should also explain the contributions of each of the participants (if you are working in a group of 2 or 3). There will also be a group presentation component.

Grading

- Class participation: 10%
  - Active in-class participation is a requirement of the course. 'Active participation' means that you should regularly ask thoughtful questions in class during lectures and tutorials, and participate with the group exercises during tutorials. If you are habitually absent from class, leave early without letting us know ahead of time, or are otherwise disengaged (e.g. on your smartphone), that will negatively affect your participation grade.
  - If you miss a class, it is up to you to borrow notes from someone, ask other students about changes to the reading/homework schedule, etc. Please don't ask me to go over what we did in class.
- Quizzes: 10%
  - Quizzes will be available via Canvas. They are designed to test your on-going knowledge of the course content, specifically the conceptual content.
- Programming assessment (approx. 3 pieces): 5%
- Homework (approx. 3 pieces): 30%
  - Assignments will be uploaded to the course website.
  - Assignments must be submitted via Canvas.

- Late assignments will NOT be accepted, except under excused circumstances.
- Emailed assignments will NOT be accepted, except under excused circumstances.
- In general, I will distribute the assignments one week before they are due, in class and/or on e-Learning.
- I will use automatic checks for overlap between your code and other students' code.
- Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade.
- Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
- Clarification: to do well in the HWs, you should examine the HW immediately, pay attention to the number of points assigned to each section, allocate a sufficient amount of time per section, and ask for clarification of anything immediately via a Canvas Discussion thread. Typically, most of the points are allocated in later sections of the HWs, therefore you should persevere and do not give up.

- Lab work (approx. 4 pieces) : 10%
  - Completion of lab exercises. You should submit the answers of the lab exercises. Generally, the deadline is a week after each of the lab sessions.
  - Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
  - Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade.
  - Clarification: to do well in the lab exercises, you should work on the lab exercises as much as you can in our lab sessions, and ask for clarification of anything during class and via a Canvas Discussion thread.
- Final project: 35%
  - Project presentations/demos: you are also required to present your work as a group. The exact format is to be determined. It will likely be a poster session open to the public.

- Submission of an individual write-up of your final project (max. 10 pages) – submitted individually.
- Submission of the group project files (the data, codes and presentation files) – submitted by your team representative.
- The project will have four milestones.
  - Project milestone 1: Create your project team (max 3 people) and propose three research proposals to be developed and submit them for review.
  - Project milestone 2: Meet with instructors to select one proposal to develop further. Submit a draft for a developed version of this proposal
  - Project milestone 3: Submit the final proposal based on feedback from the instructors
  - Project milestone 4: Present your research as a poster

Grading scale

A: 92-100, A-: 88-91.9,
B+ 85-87.9, B: 81-84.9, B-: 78-80.9,
C+: 75-77.9, C: 71-74.9, C-: 68-70.9,
D+: 65-67.9, D: 61-64.9, D-: 58-60.9,
E: Below 58

Expectations

I expect everyone to come to class and be actively engaged. I am confident that you will find it easier to master the course material by hearing it presented and also by asking questions when you don't understand something. I do not wish to see you being distracted by social media, email, and the web, therefore please avoid using your laptop, smartphone, iPad, or the like during class, except if they are needed for a class activity, such as note-taking.

Any evidence of plagiarism on problem sets will result in disciplinary penalties. In this course specifically, I expect you to do your own programming for the homework assignments. You will not learn anything if you simply copy and submit a classmate's code or code you find on the internet as your own. However, if you are stuck on a programming problem or a non-programming part of an assignment, **you are free, and indeed encouraged, to consult with your classmates (or with resources on the web) about the problems you are having**. Sharing ideas with others is extraordinarily helpful in figuring things out, and understanding a topic more deeply (both for the question asker and answerer). Once you have finished your discussions, you must write up your own code and answers, and the product you turn in should represent your work alone and not something copied from the work of your classmate. You should also note on each your assignments who or which internet resources you have consulted with. On the final project, you may work freely with the members of your group, though again I expect you to give credit to any other resources you consult.

Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies. Click here to read the university attendance

policies.

Academic honesty

You are required to abide by the Student Honor Code. Any violation of the academic integrity expected of you will result in a minimum academic sanction of failing grade on the assignment or assessment. Any alleged violations of the Student Honor Code will result in a referral to Student Conduct and Conflict Resolution. Please review the Student Honor Code and Student Conduct Code at sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/.

Students should be aware of their faculty's policy on collaboration, should understand how to properly cite sources, and should not give nor receive an improper academic advantage in any manner through any medium. No student may work or collaborate with another person on any academic activity in this course. Should group work be assigned or this class policy change, I will provide that in writing on the individual assignment instructions.

Remember you are bound by the UF Honor Pledge:

> We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity by abiding by the Student 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."

Evaluation statement

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/.

Accommodation

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center. Click here to get started with the Disability Resource Center. 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.

Health and Wellness

If you or a friend is in distress, please contact umatter@ufl.edu or 352-392-1575 so that a U Matter We Care team member can reach out to the student in distress.

Course outline

Readings should be completed **before** the class date listed.

Acronyms:

- J&M: Jurafsky and Martin (2008)
- L&C: Language and Computer (2013)
- H: Hetland (2008)
- BKL: Bird et al (2009)

Week	Topic	Reading	Assignment
W1	Introduction and Syllabus	J&M: Ch 1 (L&C Ch 1: Optional)	
	Python: Intro	LinkedIn Learning	
W2	Python: Intro	LinkedIn Learning	
W3	Python: Data types & Files	H: Ch 2-4, (Ch 11: Optional)	Programming assessment
	Python: Control & Reg Exp		
	Text Normalization	J&M: Ch 8	
W4	Lab: Text Processing	BKL: Ch 1,2,3	Programming assessment
	Edit Distance	J&M: Ch 3	
	The Noisy Channel	J&M: Ch 5, L&C: Ch2	
W5	The Noisy Channel	J&M: Ch 5, L&C: Ch 2	HW 1 set, programming assessment
	N-Grams	J&M: Ch 4	
W6	N-Grams	J&M: Ch 4	HW 1 due
	Lab: N-Grams		
W7	Machine Learning: Overview	T. Mitchel. (2017)	Project milestone 1 due
	Evaluation and Error analysis	Resnik & Lin (2010), Kummerfeld et al. (2012)	
W8	Evaluation and Error analysis	Resnik & Lin (2010), Kummerfeld et al. (2012)	HW 2 set
	Regression and Maximum Entropy	J&M: Ch 6	
W9	Regression and Maximum Entropy	J&M: Ch 6	HW 2 due, Project milestone 2 due
	Lab: Regression		
W10	Part of Speech tagging	J&M: Ch 5	
W11	Lab: Tagging	BKL: Ch 5	HW 3 Set,

Week	Topic	Reading	Assignment
			Project milestone 3 due
	Vector Semantics	TBD	
W12	Vector Semantics Lab: Semantics		HW 3 due
	Naive Bayes Classification	TBD	
W13	Sentiment Analysis Sentiment, Affect, and Connotation	TBD	HW 4 Set
W14	Lab: Sentiment Analysis		HW 4 due
	Modelling Grammar	TBD	
	Modelling Intuitions	TBD	
W15	Ethics Poster session		Project milestone 4 due

Full references:

Bird, Steven, Ewan Klein, and Edward Loper (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Hetland, Magnus Lie (2008). Beginning Python from Novice to Professional. Wiley.

Jurafsky, Daniel and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. Isbn: 0131873210.

Mitchell, T.M.: Key Ideas in Machine Learning. http://www.cs.cmu.edu/%7Etom/mlbook/keyIdeas.pdf (2017). Accessed 28 Jan 2019

Resnik, P. & J. Lin (2010), Evaluation of nlp systems, in A. Clark, C. Fox, & S. Lappin (eds.), Handbook of Computational Linguistics and Natural Language Processing, Wiley-Blackwell, Oxford.

Kummerfeld, J. K., Hall, D., Curran, J. R., & Klein, D. (2012, July). Parser showdown at the wall street corral: An empirical investigation of error types in parser output. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1048-1059).

LIN5XXX Introduction to Computational Linguistics Instructor: Dr. Kevin Tang

General Information

Course description

Computational linguistics is the study of natural language from a computational perspective. It encompasses both applied (engineering) and theoretical (cognitive) issues, ranging from speech and language technology to formal aspects of theoretical linguistic models.

This course is for YOU if you have ever wondered: How your phone can do predictive texting? How does a computer understand the syntax of a sentence? How Amazon can process billions of reviews? How to automatically process large corpora (big data)? How to discover linguistic structures from language data automatically? How to model our linguistic intuitions of grammaticality?

This course is for YOU if you want to build a foundation for areas such as data science, the tech industry https://www.linguisticsociety.org/resource/webinar-linguists-and-linguistics-tech or for applying for a Master's in Computational Linguistics.

This course is for YOU if you want to create a computational linguistics project like these ones by last year's students: https://slam.lin.ufl.edu/2020/04/20/intro2compling-symposium-spring2020/.

In this class, we will survey various topics and tasks in computational linguistics. While we will cover some of the basics of Natural Language Processing (which we will consider a separate subfield), this class will not focus on one specific approach (such as deep learning). Students in this class are expected to have a background in either computer science or linguistics, but not necessarily both. Expect this class to be difficult at times and easy at others. We hope to offer something new and interesting for everyone.

Please also check out this page to learn more about computational linguistics at UF: https://slam.lin.ufl.edu/computational-linguistics/

Objectives

On completion of this course, you should:

- Be able to use computational linguistic tools and resources, and be able to describe how they are applied in research in both computational linguistics and other subfields

- Have a rough sense of the state of the art in this subfield
- Be able to conceptualize problems from the perspective of computational linguistics
- Be able to code in Python and use JyputerNotebook.

Time and place

WHERE: TBA

WHEN: TBA

Instructor information

Instructor: Dr. Kevin Tang

– https://slam.lin.ufl.edu/

Email:

- `tang.kevin@ufl.edu`

Offices:

- 4017 Turlington Hall, Gainesville, FL 32611-5454

Office hours:

- TBD or by appointment.
- For information on what are office hours and how to make use of them properly see: http://lsc.cornell.edu/wp-content/uploads/2015/10/What-Are-Office-Hours.pdf.

Requirements

Prerequisites

None

Programming training

Programming is not the focus of this course, but knowing how to program is an essential skill needed to do computational linguistics. I have engineered a number of ways to get you all up to speed with programming.

- **The first three weeks** will cover a brief introduction to programming in Python, with a particular focus on learning how to process written text, and JyputerNotebook (an interactive computational environment).
- **Two free online courses** will be assigned as assignments in a flipped class room format.
- After these first few weeks, you should have **a basic foundation of Python** to be able to tackle non-coding related course components.
- The course will include a complementary survey of the **Python Natural Language Toolkit (NLTK)** which lowers the need of coding for certain topics of the course.

Note that the programming training will not be entirely comprehensive as one can never finish

learning to code. You need to be prepared to be **highly motivated** when it comes to learning how to code using Python. Learning to code is *not* like learning a language or learning to do math (see https://news.mit.edu/2020/brain-reading-computer-code-1215). This means you should complete all the assigned online courses and exercises, and make use of the class time, Canvas discussion forums, office hours, and any supplementary resources provided (e.g., the Python textbook).

To stay motivated, it is important to consider why should you learn to code and how coding can improve your career prospect and open doors. Here are very cool animated videos on the basic concepts of coding by Karlie Kloss and TedTalk-Ed.

- "Coding is a superpower": https://www.youtube.com/watch?v=Bwiln7v0fdc
- "Karlie Kloss Explains How Computers Work": https://www.youtube.com/watch?v=_zeV-xWSJoU
- "Variables, Functions, and Arrays with Karlie Kloss": https://www.youtube.com/watch?v=MXkGONNYKHo
- "Think like a coder" (10 episodes): https://youtu.be/KFVdHDMcepw

To understand your suitability, I encourage you to check out the two free online courses that we will assign as assignments during the first three weeks. As always, you can always **reach out to me at** `tang.kevin@ufl.edu` **to discuss your suitability**.

- Note: As a UF student, the courses below are free (see instructions here: https://elearning.ufl.edu/supported-services/linkedin-learning/)
- Course 1: Python Essential Training (estimated duration: 4h 45m) by Bill Weinman https://www.linkedin.com/learning/python-essential-training-2
- Course 2: Introducing Jupyter by Josh McQuiston (estimated duration: 53m) https://www.linkedin.com/learning/introducing-jupyter/

Course website

- There is a course website on Canvas
- Let me know if you are not on the site.

Textbook and Reading list

Main textbook (Second edition):

- Jurafsky, Daniel and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. Isbn: 0131873210.
- Make sure you get the second edition! Within the second edition, the hardcover “US edition" is preferred, but you can also buy the “international edition" (which might be cheaper). There’s also an ebook version of the second edition (which again might be cheaper). They differ mostly only in a few exercises at the end of each chapter.

Supplementary textbook (Freely available online):

- Freely available online: https://www.nltk.org/book/

- Dickinson, Brew, and Meuers (2013), *Language and Computers*
- Available for reading online through the UF library (and download up to 51 pages)

Python textbook (Available at UF):

- Available for download as an ebook through the UF library

Programming Web Resources (Available at UF):

- Note: As a UF student, the LinkedIn courses below are free (see instructions here: https://elearning.ufl.edu/supported-services/linkedin-learning/)
- Course 1: Python Essential Training (estimated duration: 4h 45m) by Bill Weinman https://www.linkedin.com/learning/python-essential-training-2
- Course 2: Introducing Jupyter by Josh McQuiston (estimated duration: 53m) https://www.linkedin.com/learning/introducing-jupyter/
- Course 3 (Optional, but recommended): https://www.linkedin.com/learning/python-functions-for-data-science/
- Additionally, there are numerous online resources for Python, and you may find them a convenient supplement to the Hetland book, especially if you're an experienced programmer. https://docs.python.org/3/tutorial/, and https://www.codecademy.com/learn/learn-python.
- If you have a specific problem or question, Stack Overflow is a good place to look for answers: stackoverflow.com. I recommend searching the site for your question before asking it yourself.
- Learning how to google specific questions to debug is highly recommended (http://letmegooglethat.com/)
- Readings from the textbooks will be supplemented by other readings and materials throughout the semester (made available on the website for the course).

Course requirements

1. **Lectures and Reading:** The content of the course will be presented through lectures and reading assignments. You should attend lectures and complete the reading assignment by the date of the lecture with which the reading is associated. The material in the lectures and readings will not be identical, so you will need to both attend lectures and do the readings to succeed. Homework and quizzes will presume familiarity with both the readings and the lectures.
2. **Quizzes:** There will be a number of short quizzes in this course which are designed to test your conceptual knowledge of computational linguistics.
3. **Programming assessments**: There will be a number of programming assignments.
4. **Lab exercises**: There will be a number of lab sessions, indicated in the schedule below, which will allow you to practice the skills you have learned in the course in a supervised environment. They will test your knowledge of the material we discuss in class – a combination of programming and conceptual knowledge. **You are encouraged to work**

in a group with these lab exercises. In my experience, students benefit a lot from working with each other.

5. **Homework assignments:** There will be a number of homework assignments (on the order of 3 or 4) in this course, which will involve programming. **You are encouraged to work in a group with these lab exercises.** These assignments will require you to *implement algorithms* from computational linguistics, t*est them on data sets*, and suggest and *explore potential improvements*.
6. **Final Project**: **You will work in a group of 2-3 people on a final project** that builds in some way on the material we cover in class and/or connects with related literature. The amount of work that I expect from you will, of course, vary by the number of people who contribute, but you may find it helpful and interesting to work with others who have backgrounds different from yours. The range of allowable topics is very flexible: it may relate to a personal passion of your (e.g., comic books) or to an area of research. I would encourage you to try to find a topic that inspires you (and your partners), since this is something that you will be spending a good deal of time on.

Once you have identified a topic with your group, you should schedule an appointment with me to finalize the idea. Your project must include a computational implementation of some sort, and must be presented in a written report that explains what you have done, how it relates to past work, and what you have learned from your results. You should also explain the contributions of each of the participants (if you are working in a group of 2 or 3). There will also be a group presentation component.

Grading

- Class participation: 10%
  - Active in-class participation is a requiredsment of the course. 'Active participation' means that you should regularly ask thoughtful questions in class during lectures and tutorials, and participate with the group exercises during tutorials. If you are habitually absent from class, leave early without letting us know ahead of time, or are otherwise disengaged (e.g. on your smartphone), that will negatively affect your participation grade.
  - If you miss a class, it is up to you to borrow notes from someone, ask other students about changes to the reading/homework schedule, etc. Please don't ask me to go over what we did in class.
- Quizzes: 10%
  - Quizzes will be available via Canvas. They are designed to test your on-going knowledge of the course content, specificially the conceptual content.
- Programming assessment (approx. 3 pieces): 5%
- Homework (approx. 2 pieces): 20%
  - Assignments will be uploaded to the course website.
  - Assignments must be submitted via Canvas.

- Late assignments will NOT be accepted, except under excused circumstances.
- Emailed assignments will NOT be accepted, except under excused circumstances.
- In general, I will distribute the assignments one week before they are due, in class and/or on e-Learning.
- I will use automatic checks for overlap between your code and other students' code.
- Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade.
- Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
- Clarification: to do well in the HWs, you should examine the HW immediately, pay attention to the number of points assigned to each section, allocate a sufficient amount of time per section, and ask for clarification of anything immediately via a Canvas Discussion thread. Typically, most of the points are allocated in later sections of the HWs, therefore you should persevere and do not give up.

- Lab work (approx. 4 pieces) : 10%
  - Completion of lab exercises. You should submit the answers of the lab exercises. Generally, the deadline is a week after each of the lab sessions.
  - Submit three files, unless mentioned otherwise (.ipynb (JyputerNotebook), .html (the output of JyputerNotebook) and .pdf (a pdf rendered version of the html). I will need the .pdf file for grading purposes (giving you comments directly in the document). If you do not submit all three files, I will not be able to grade properly and this will negatively affect your grade. Files submitted using the 'comment' feature will not be accepted.
  - Length restrictions: be careful about your .html and .pdf files in terms of what is actually printed. If you should accidentally generate an excessive number of pages (e.g., you print out the whole dataset to inspect it, leading to a 300 page pdf file), I will not be able to grade properly and this will negatively affect your grade.
  - Clarification: to do well in the lab exercises, you should work on the lab exercises as much as you can in our lab sessions, and ask for clarification of anything during class and via a Canvas Discussion thread.
- Final project: 45%
  - Project presentations/demos: you are also required to present your work as a group. The exact format is to be determined. It will likely be a poster session open to the public.

- Submission of an individual write-up of your final project (max. 10 pages) – submitted individually.
- Submission of the group project files (the data, codes and presentation files) – submitted by your team representative.
- Select a computational linguistics conference venue as a target outlet and prepare as a group a conference paper (typically 7 pages, two columns) as well as an abstract for submission
- The project will have four milestones.
  - Project milestone 1: Create your project team (max 3 people) and propose three research proposals to be developed and submit them for review.
  - Project milestone 2: Meet with instructors to select one proposal to develop further. Submit a draft for a developed version of this proposal
  - Project milestone 3: Submit the final proposal based on feedback from the instructors
  - Project milestone 4: Present your research as a poster

Grading scale

A: 92-100, A-: 88-91.9,
B+ 85-87.9, B: 81-84.9, B-: 78-80.9,
C+: 75-77.9, C: 71-74.9, C-: 68-70.9,
D+: 65-67.9, D: 61-64.9, D-: 58-60.9,
E: Below 58

Expectations

I expect everyone to come to class and be actively engaged. I am confident that you will find it easier to master the course material by hearing it presented and also by asking questions when you don't understand something. I do not wish to see you being distracted by social media, email, and the web, therefore please avoid using your laptop, smartphone, iPad, or the like during class, except if they are needed for a class activity, such as note-taking.

Any evidence of plagiarism on problem sets will result in disciplinary penalties. In this course specifically, I expect you to do your own programming for the homework assignments. You will not learn anything if you simply copy and submit a classmate's code or code you find on the internet as your own. However, if you are stuck on a programming problem or a non-programming part of an assignment, **you are free, and indeed encouraged, to consult with your classmates (or with resources on the web) about the problems you are having**. Sharing ideas with others is extraordinarily helpful in figuring things out, and understanding a topic more deeply (both for the question asker and answerer). Once you have finished your discussions, you must write up your own code and answers, and the product you turn in should represent your work alone and not something copied from the work of your classmate. You should also note on each your assignments who or which internet resources you have consulted with. On the final project, you may work freely with the members of your group, though again I expect you to give

credit to any other resources you consult.

Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies. Click here to read the university attendance policies.

Academic honesty

You are required to abide by the Student Honor Code. Any violation of the academic integrity expected of you will result in a minimum academic sanction of failing grade on the assignment or assessment. Any alleged violations of the Student Honor Code will result in a referral to Student Conduct and Conflict Resolution. Please review the Student Honor Code and Student Conduct Code at sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/.

Students should be aware of their faculty's policy on collaboration, should understand how to properly cite sources, and should not give nor receive an improper academic advantage in any manner through any medium. No student may work or collaborate with another person on any academic activity in this course. Should group work be assigned or this class policy change, I will provide that in writing on the individual assignment instructions.

Remember you are bound by the UF Honor Pledge:

> We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity by abiding by the Student 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."

Evaluation statement

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/.

Accommodation

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center. Click here to get started with the Disability Resource Center. 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.

Health and Wellness

If you or a friend is in distress, please contact umatter@ufl.edu or 352-392-1575 so that a U Matter We Care team member can reach out to the student in distress.

Course outline

Readings should be completed **before** the class date listed.

Acronyms:

- J&M: Jurafsky and Martin (2008)
- L&C: Language and Computer (2013)
- H: Hetland (2008)
- BKL: Bird et al (2009)

Week	Topic	Reading	Assignment
W1	Introduction and Syllabus	J&M: Ch 1 (L&C Ch 1: Optional)	
	Python: Intro	LinkedIn Learning	
W2	Python: Intro	LinkedIn Learning	
W3	Python: Data types & Files	H: Ch 2-4, (Ch 11: Optional)	Programming assessment
	Python: Control & Reg Exp		
	Text Normalization	J&M: Ch 8	
W4	Lab: Text Processing	BKL: Ch 1,2,3	Programming assessment
	Edit Distance	J&M: Ch 3	
	The Noisy Channel	J&M: Ch 5, L&C: Ch2	
W5	The Noisy Channel	J&M: Ch 5, L&C: Ch 2	HW 1 set, programming assessment
	N-Grams	J&M: Ch 4	
W6	N-Grams	J&M: Ch 4	HW 1 due
	Lab: N-Grams		
W7	Machine Learning: Overview	T. Mitchel. (2017)	Project milestone 1 due
	Evaluation and Error analysis	Resnik & Lin (2010), Kummerfeld et al. (2012)	
W8	Evaluation and Error analysis	Resnik & Lin (2010), Kummerfeld et al. (2012)	HW 2 set
	Regression and Maximum Entropy	J&M: Ch 6	
W9	Regression and Maximum Entropy	J&M: Ch 6	HW 2 due, Project milestone 2 due
	Lab: Regression		
W10	Part of Speech tagging	J&M: Ch 5	
W11	Lab: Tagging	BKL: Ch 5	HW 3 Set,

Week	Topic	Reading	Assignment
			Project milestone 3 due
	Vector Semantics	TBD	
W12	Vector Semantics Lab: Semantics		HW 3 due
	Naive Bayes Classification	TBD	
W13	Sentiment Analysis Sentiment, Affect, and Connotation	TBD	HW 4 Set
W14	Lab: Sentiment Analysis		HW 4 due
	Modelling Grammar	TBD	
	Modelling Intuitions	TBD	
W15	Ethics Poster session		Project milestone 4 due

Full references:

Bird, Steven, Ewan Klein, and Edward Loper (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Hetland, Magnus Lie (2008). Beginning Python from Novice to Professional. Wiley.

Jurafsky, Daniel and James H. Martin (Sept. 2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd edition). Prentice Hall. Isbn: 0131873210.

Mitchell, T.M.: Key Ideas in Machine Learning. http://www.cs.cmu.edu/%7Etom/mlbook/keyIdeas.pdf (2017). Accessed 28 Jan 2019

Resnik, P. & J. Lin (2010), Evaluation of nlp systems, in A. Clark, C. Fox, & S. Lappin (eds.), Handbook of Computational Linguistics and Natural Language Processing, Wiley-Blackwell, Oxford.

Kummerfeld, J. K., Hall, D., Curran, J. R., & Klein, D. (2012, July). Parser showdown at the wall street corral: An empirical investigation of error types in parser output. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1048-1059).