Lin511-001 - Computational Linguistics

Josef Fruehwald

January 2024

Sections

1 Key info

Key Info

Where:	Whitehall, Rm 205
When:	Tues & Thurs, 09:30 - 10:45
Prereq:	Lin221
Credits:	3

Instructor

	Dr. Josef Fruehwald
email:	josef.fruehwald@uky.edu
office hours:	Mons, 12:00pm - 02:00pm
office hours location:	Breckinridge, Rm 10

2 Course at a Glance

What you'll learn: Computational approaches to linguistic analysis; Computational tools (python, regular expressions, Git, GitHub).

What you'll do: In class exercises; Programming assignments.

What you'll need: A computer with a physical keyboard; A GitHub account.

The final-est deadline April 30, 2024

Attendance Policy Attendance is crucial for successful completion of the course, but there are no grade penalties.

Late Work Policy 2 day penalty free grace period on all assignments, 5% flat penalty afterwards. See Late Submissions and Re-submissions

3 Course Description

There are two important components to this course

1. This is an introduction to computational linguistics, with an emphasis on **linguistics**. We'll be learning about approaches to computation as it relates to linguistic theory (e.g. phonological rules, syntactic parsing, etc) as well as computation involved in processing linguistic data (e.g. Large Language Models, Speech-to-Text etc.).
2. This will be an introduction to some practical aspects of general purpose *computation*, including basics of file system organization, version control, Integrated Development Environments, command-line interfaces, and program writing (specifically in Python).

Have you ever said one these things?

- "My computer hates me."
- "I'm not a tech person."

As part of our course meetings, I'll be labelling these and other similar statements as "negative self-talk". Instead, I'll encourage you to try different statements, like

- "I'm not familiar with these concepts yet."
- "Up to now, I've found these methods opaque."

We can acknowledge your current struggle or confusion with computation or technology, while also acknowledging that their use is a skill, not a talent, and that skill can be built upon and improved with experience and *practice.*

---

4 Learning Outcomes

After attending class meetings and completion of the coursework, students should be able to

- Describe the computational methods used to model and process linguistic structures.
- Use Regular Expressions to search and match strings.
- Write a python program to linguistically parse language data.
- Critically evaluate claims made about modern natural language processing applications.

5 Course Materials

We will be using a mixture of textbooks and online resources for the course. These are currently available for free online. The labels **SLP**, **NLTK** and **TP** will be used to refer to each book in the reading schedule.

SLP: Daniel Jurafsky and James H Martin. 202\d. *Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics, and Speech Recognition* Third Edition.

NLTK: Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*

TP: Allen B. Downey. 2015. *Think Python,* 2nd edition

6 Course Technology

As course with a practical programming component, we're going to be using a number of technical programs. There will be specific class time set aside for setting these up.

Command-Line Interfaces

Command-line Interfaces (CLI) are text-only ways to interact with your computer, including accessing its files and running programs. We will all be accessing the CLI through Visual Studio Code. For Windows users, this will also involve installing Windows Subsystem for Linux.

Python Scripts & Jupyter Notebooks

Python is the programming language that we'll be using to do most of our work and analysis in this course. We'll also be using extensions to the Python programming language in the form of freely available Python libraries, such as nltk and numpy.

We'll be writing both *python scripts* as well as using *Jupyter notebooks* to interact with python.

Visual Studio Code

The fact we will be "writing" python scripts implies we will be writing them *in* something. The program we will be using for composing our programs will be Visual Studio Code. VS Code is a general purpose Integrated Development Environment (IDE).

As "formal languages", programming languages are very sensitive to any kind of typo or formatting error. The purpose IDEs is to provide you support to avoid these typos & errors in the first place, and to warn you when they exist.

Git/Github

Git is a "Version Control System" that lets you keep track of changes on software projects. Github is a service that allows online hosting of Git projects. You will need to create a free a Github account for the course.

There will be a number of course assignments that you will submit via commits to GitHub.

Canvas

Canvas will be used to make course announcements, and to distribute assignment links.

---

7 Communications

I will respond to emails in a timely manner during normal working hours, but it may take longer if you email me after 5pm on weekdays, or any time during the weekend.

---

8 Course Schedule

The topics and readings listed here are the tentative schedule for the course. We may find, in the room, that some topics will take longer than initially scheduled.

Week 1

Dates: Jan 8-12

Topics: Setup, What is Computational Linguistics?

Alternative Office Hours:

Friday, Jan 12	12pm to 2pm	LinCoLab (Breckinridge, Room10)

Readings

Follow posted tutorials for setting up our course technology.

Week 2

Dates: Jan 15-19

Topics: Files, Text, Strings, Regular Expressions, Finite State Automata

No Office Hours for MLK Jr Day.

Readings:

- [Course Notes: What *is* Python](#)
- [SLP Chapter 2, section 2.1 (pdf)](#)
- [TP Chapter 1](#)

Week 3

Dates: Jan 22-26

Topics: Text Normalization, Tokenization, Python Variables, Python Functions

Readings

- [SLP Chapter 2, sections 2.2, 2.3, 2.4](#)
- [TP, Chapter 2](#)
- [TP Chapter 3](#)
- [NLTK Book, Chapter 3](#)

Week 4

Dates: Jan 29 - Feb 02

Topics: More

Week 5

Dates: Feb 05-09

Topics: More

Week 6

Dates: Feb 12 - 16

Topics: More

Week 7

Dates: Feb 19-23

Topics: More

Week 8

Dates: Feb 26 - Mar 01

Topics: More

Week 9

Dates: Mar 04-08

Topics: More

Spring Break

Week 10

Dates: Mar 18-22

Topics: More

Week 11

Dates: Mar 25-29

Topics: More

Week 12

Dates: Apr 01-05

Topics: More

Week 13

Dates: Apr 08-12

Topics: More

Week 14

Dates: Apr 15-19

Topics: More

Week 15

Dates: Apr 22-26

Topics: More

9 Course Evaluation

Grade Components

Weekly Exercises	45%
In-Class Exercises	25%
Final Project	20%
Engagement	10%

Grading Scale

A	>= 90
B	80 to 89
C	70 to 79
D	60 to 69
E	<= 59

Assignment Submission

Some assignments will be created in GitHub Classroom, and their invite code will be posted to canvas. Some of these assignments will have "autograding" tests enabled. These "autogrades" are intended to be feedback to help you fine tune your code, and **are not** meant to be the final grade you will get for the assignment. Only grades as they are appear on Canvas are your official grade.

Final Project

A final capstone project for the course. This could be a report, an extension of an earlier exercise, or some other agreed upon format.

Engagement

Inspired by Kirby Conrod's approach to Participation Grades

> This portion of the grade is a way for me to give you credit for informal/unstructured collaborative work that you do. Participation and collaboration are strong predictors of success and learning retention, so please make an effort to find a way that works well for you to participate and engage with your colleagues.

A well known process for solving programming problems is "Rubber Duck Debugging." It works by describing how each step of a program is supposed to work to another person or, as the name suggests, a rubber duck. Often the solution to the problem or the typo causing the bug jumps out at you during the process. Having a study buddy or study group could be really helpful if only for this purpose.

---

10 Late Submissions and Re-submissions

Every graded piece of work will have a due date. After a 2 day grace period, there will be a single, flat 5% deduction from late work, whenever it is submitted between the due date and the The Final-est Deadline

Midterm Grades

I will submit midterm grades on March 08, 2024, at the end of the midterm grading window. Any unsubmitted assignments that were due before March 08 will be given a grade of 0, **_BUT_** you can still submit those assignments after March 08 for their inclusion in the final grade.

The Final-est Deadline

The final-est deadline by which to submit any material to be graded is **April 30, 2024**. I have to set this hard deadline in order to have enough time to conclude final grading in time for the university's final grade submission deadline.

---

11 Group Work and Code Sources

It is acceptable to collaborate and confer with other students in the course. Any collaboration should be indicated in the assignment submission. You may also refer to code sources from elsewhere on the internet, as long as you also document the source, and explain what the code does. You might not receive credit for code which has been copied wholesale from another online source or from another student without credit or documentation.

Large Language Model (a.k.a. AI) Generated Code

There are a number of services that will generate code based on natural language queries. Some words of warning:

Fluent BS

Large Language Models have been found to generate code that looks superficially correct, but often does not actually run properly, or do what the human asker wanted. Being able to successfully identify where or why code does not work correctly is not always straight forward. This issue led the Q&A site StackOverflow to ban submissions generated by LLMs, stating

> [...] because GPT is good enough to convince users of the site that the answer holds merit, signals the community typically use to determine the legitimacy of their peers' contributions frequently fail to detect severe issues with GPT-generated answers.

Explain what the code does

As stated above, you should provide credit to any external sources you turned to for code help, and explain what the resulting code does.

---

12 Attendance and Engagement

You are expected to attend all scheduled course meetings. It would be helpful, but not necessary, if you let me know in advance if you are going to miss any lectures.

If you feel sick in any way, including but not limited to the well-known symptoms of COVID-19 (loss of taste or smell, a new and persistent cough, high fever, etc), do not come to class. There are other mechanisms for demonstrating engagement than attending lectures.

I will also expect all of us in the course to treat each other with respect and civility in all aspects of the course, including

- In the audio of a Zoom meeting
- In the text chat of a Zoom meeting
- On any course discussion boards or other forums.

---

13 Academic Conduct

UK Senate rules on academic offences

Appropriating someone else's work and portraying it as your own is cheating. Collaborating with someone and portraying that work as solely your own is cheating. Obtaining answers to homework assignments or exams from previous semesters is cheating. Using an internet search engine to look up a question and reporting that answer as your own is cheating. Falsifying data or experimental results is cheating. If you are unsure about whether a specific action is cheating, you may check with me.

The minimum penalty for a first offense is a zero on the assignment on which the offense occurred. If the offense is considered severe or if the student has other academic offenses on their record, more serious penalties, up to suspension from the University may be imposed.

When students submit work purporting to be their own, but which in any way borrows ideas, organization, wording or anything else from another source without appropriate acknowledgement of the fact, the students are guilty of plagiarism. Plagiarism includes reproducing someone else's work, whether it be a published article, chapter of a book, a paper from a friend or some file, or something similar to this. Plagiarism also includes the practice of employing or allowing another person to alter or revise the work which a student submits as their own, whoever that other person may be.

Students may discuss assignments among themselves or with an instructor or tutor, but when the actual work is done, it must be done by the student, and the student alone. When a student's assignment involves research in outside sources of information, the student must carefully acknowledge exactly what, where and how they employed them. If the words of someone else are used, the student must put quotation marks around the passage in question and add an appropriate indication of its origin. Making simple changes while leaving the organization, content and phraseology intact is plagiaristic. However, nothing in these Rules shall apply to those ideas which are so generally and freely circulated as to be a part of the public domain (University Senate Rules Section 6.3.1).

---

14 University Academic Policy Statements

Link to University Senate Academic Policy Statements

Excused Absences and Acceptable Excuses

Excused Absences: *Senate Rules 5.2.5.2.1* defines the following as acceptable reasons for excused absences: (a) significant illness, (b) death of a family member, (c) trips for members of student organizations sponsored by an educational unit, trips for University classes, and trips for participation in intercollegiate athletic events, (d) major religious holidays, (e) interviews for graduate/professional school or full-time employment post-graduation, and (f) other circumstances found to fit "reasonable cause for nonattendance" by the instructor of record. Students should notify the professor of absences prior to class when possible.

If a course syllabus requires specific interactions (e.g., with the instructor or other students), in situations where a student's total EXCUSED absences exceed 1/5 (or 20%) of the required interactions for the course, the student shall have the right to request and receive a "W," or the Instructor of Record may award an "I" for the course if the student declines a "W." (*Senate Rules 5.2.5.2.3.1*)

Religious Observances

Religious Observances: Students anticipating an absence for a major religious holiday are responsible for notifying the instructor in writing of anticipated absences due to their observance of such holidays. *Senate Rules 5.2.5.2.1(4) requires faculty to include any notification requirements within the syllabus. If no requirement is specified, two weeks prior to the absence is reasonable and should not be given any later. Information regarding major religious holidays may be obtained through* [the Ombud's website](#)*or calling 859-257-3737.*

Verification of Absences

Verification of Absences:Students may be asked to verify their absences in order for them to be considered excused. *Senate Rule 5.2.5.2.1* states that faculty have the right to request appropriate verification when students claim an excused absence due to: significant illness; death in the household, trips for classes, trips sponsored by an educational unit and trips for participation related to intercollegiate athletic events; and interviews for full-time job opportunities after graduation and interviews for graduate and professional school. (Appropriate notification of absences due to University-related trips is required prior to the absence when feasible and in no case more than one week after the absence.)

Make-Up Work

Make-Up Work: Students missing any graded work due to an excused absence are responsible: for informing the Instructor of Record about their excused absence within one week following the period of the excused absence (except where prior notification is required); and for making up the missed work. The instructor must give the student an opportunity to make up the work and/or the exams missed due to the excused absence, and shall do so, if feasible, during the semester in which the absence occurred. The instructor shall provide the student with an opportunity to make up the graded work and may not simply calculate the student's grade on the basis of the other course requirements, unless the student agrees in writing. According to *SR 5.2.5.2.2*, if a student adds a class after the first day of classes and misses graded work, the instructor must provide the student with an opportunity to make up any graded work.

Excused Absences for Military Duties

Excused Absences for Military Duties: If a student is required to be absent for one-fifth or less of the required course interactions (e.g., class meetings) due to military duties, the following procedure (per *SR 5.2.5.2.3.2*) shall apply:

1. Once a student is aware of a call to duty, the student shall provide a copy of the military orders to the Director of the Veterans Resource Center. The student shall also provide the Director with a list of his/her courses and instructors.

2. The Director will verify the orders with the appropriate military authority, and on behalf of the military student, notify each Instructor of Record via Department Letterhead as to the known extent of the absence.

3. The Instructor of Record shall not penalize the student's absence in any way and shall provide accommodations and timeframes so that the student can make up missed assignments, quizzes, and tests in a mutually agreed upon manner.

Unexcused Absences

Unexcused Absences: If an attendance/interaction policy is not stated in the course syllabus or the policy does not include a penalty to the student, the instructor cannot penalize a student for any unexcused absences. (*SR 5.2.5.2.3.3*)

Prep Week and Reading Days

Prep Week and Reading Days: Per *Senate Rules 5.2.5.6*, the last week of instruction of a regular semester is termed "Prep Week." This phrase also refers to the last three days of instruction of the summer session and winter intersession. The Prep Week rule applies to ALL courses taught in the fall semester, spring semester, and summer session, including those taught by distance learning or in a format that has been compressed into less than one semester or session. This rule does not apply to courses in professional programs in colleges that have University Senate approval to have their own calendar.

Make-up exams and quizzes are allowed during Prep Week. In cases of "Take Home" final examinations, students shall not be required to return the completed examination before the regularly scheduled examination period for that course. No written examinations, including final examinations, may be scheduled during the Prep Week. No quizzes may be given during Prep Week. No project/lab practicals/paper/presentation deadlines or oral/listening examinations may fall during the Prep Week unless it was scheduled in the syllabus AND the course has no final examination (or assignment that acts as a final examination) scheduled during finals week. (A course with a lab component may schedule the lab practical of the course during Prep Week if the lab portion does not also require a Final Examination during finals week.) Class participation and attendance grades are permitted during Prep Week. The *Senate Rules* permit continuing into Prep Week regularly assigned graded homework that was announced in the class syllabus.

For fall and spring semester, the Thursday and Friday of Prep Week are study days (i.e. "Reading Days"). There cannot be any required "interactions" on a Reading Day. "Interactions"

include participation in an in-class or online discussion, attendance at a guest lecture, or uploading an assignment. See *Senate Rules 9.1* for a more complete description of required interactions.

Accommodations Due to Disability

Accommodations Due to Disability: In accordance with federal law, if you have a documented disability that requires academic accommodations, please inform your instructor as soon as possible during scheduled office hours. In order to receive accommodations in a course, you must provide your instructor with a Letter of Accommodation from the Disability Resource Center (DRC). The DRC coordinates campus disability services available to students with disabilities. It is located on the corner of Rose Street and Huguelet Drive in the Multidisciplinary Science Building, Suite 407. You can reach them via phone at (859) 257-2754, via email (drc@uky.edu) or visit their website (uky.edu/DisabilityResourceCenter). DRC accommodations are not retroactive and should therefore be established with the DRC as early in the semester as is feasible.

Non-Discrimination Statement and Title IX Information

Non-discrimination and Title IX policy: In accordance with federal law, UK is committed to providing a safe learning, living, and working environment for all members of the University community. The University maintains a comprehensive program which protects all members from discrimination, harassment, and sexual misconduct. For complete information about UK's prohibition on discrimination and harassment on aspects such as race, color, ethnic origin, national origin, creed, religion, political belief, sex, and sexual orientation, please see the electronic version of *UK's Administrative Regulation 6:1* ("Policy on Discrimination and Harassment") (https://www.uky.edu/regs/ar6-1). In accordance with Title IX of the Education Amendments of 1972, the University prohibits discrimination and harassment on the basis of sex in academics, employment, and all of its programs and activities. Sexual misconduct is a form of sexual harassment in which one act is severe enough to create a hostile environment based on sex and is prohibited between members of the University community and shall not be tolerated. For more details, please see the electronic version of *Administrative Regulations 6:2* ("Policy and Procedures for Addressing and Resolving Allegations of Sexual Harassment Under Title IX and Other Forms of Sexual Misconduct") (https://www.uky.edu/regs/sites/www.uky.edu.regs/files/files/ar/ar_6.2-in...). Complaints regarding violations of University policies on discrimination, harassment, and sexual misconduct are handled by the Office of Institutional Equity and Equal Opportunity (Institutional Equity), which is located in 13 Main Building and can be reached by phone at (859) 257-8927. You can also visit Institutional Equity's website (https://www.uky.edu/eeo).

Faculty members are obligated to forward any report made by a student related to discrimination, harassment, and sexual misconduct to the Office of Institutional Equity. Students can confidentially report alleged incidences through the Violence Intervention and Prevention Center (https://www.uky.edu/vipcenter), Counseling Center (https://www.uky.edu/counselingcenter), or University Health Service (https://ukhealthcare.uky.edu/university-health-service/student-health).

Reports of discrimination, harassment, or sexual misconduct may be made via the Institutional Equity's website (https://www.uky.edu/eeo); at that site, click on "Make a Report" on the left-hand side of the page.

Regular and Substantive Interaction

Regular and Substantive Interaction: All credit-bearing courses must support regular and substantive interaction (RSI) between the students and the instructor, regardless of the course's delivery mode (e.g., in-person, hybrid, or online). Courses satisfy this requirement when course participants meet regularly as prescribed in SR 10.6, and the Instructor of Record substantively interacts with students in at least two of the following ways: provides direct instruction; assesses students' learning; provides information or responds to students' questions; and facilitates student discussions. Some exceptions allowed as per SACSCOC. For further information about the RSI requirement, see the Compliance Resources link on the Teaching, Learning and Academic Innovation Compliance page.

---

15 Diversity, Equity and Inclusion

UK Senate DEI Statement

The University of Kentucky is committed to our core values of diversity and inclusion, mutual respect and human dignity, and a sense of community (Governing Regulations XIV). We acknowledge and respect the seen and unseen diverse identities and experiences of all members of the university community (https://www.uky.edu/regs/gr14). These identities include but are not limited to those based on race, ethnicity, gender identity and expressions, ideas and perspectives, religious and cultural beliefs, sexual orientation, national origin, age, ability, and socioeconomic status. We are committed to equity and justice and providing a learning and engaging community in which every member is engaged, heard, and valued.

We strive to rectify and change behavior that is inconsistent with our principles and commitment to diversity, equity, and inclusion. If students encounter such behavior in a course, they are encouraged to speak with the instructor of record and/or the Office of Institutional Equity and Equal Opportunity. Students may also contact a faculty member within the department,

program director, the director of undergraduate or graduate studies, the department chair, any college administrator, or the dean. All of these individuals are mandatory reporters under University policies.