LING 529
Human Language Technology I

Eric Jackson

Last Revised August 19, 2022

1 Course Information

This class serves as an introduction to human language technology (HLT), an emerging interdisciplinary field that encompasses most subdisciplines of linguistics, as well as computational linguistics, natural language processing, computer science, artificial intelligence, psychology, philosophy, mathematics, and statistics. Content includes a combination of theoretical and applied topics such as (but not limited to) tokenization across languages, n-grams, word representations, basic probability theory, introductory programming, and version control.

Course Objectives

In this course, we will . . .

- cover fundamental concepts related to human language technology, such as . . .
  - tokens and their attributes
  - text normalization techniques
  - tokenization and regular expressions
  - character and word n-grams
  - basics of probability theory
  - representing words and documents as vectors
  - vector-based comparisons
- foster technical skills, such as . . .
  - linux command line basics
  - virtualization and Containerization technologies
  - version control (ex. `git`) and the feature branch workflow

Expected Learning Outcomes

Successful students will be able to . . .

- write simple programs ≤ 50 lines of code[1]
- use a version control system to manage repositories[1]
- set up a Linux-based development environment
- calculate the similarity of pairs of vectors[2]
- build a simple tokenizer[2]
- represent words and documents as vectors[2]
- estimate the probability of a word in a corpus[2]

Credits: 3 units

Prerequisites

- None

---

[1]Relates to Linguistics Department's HLT Program Outcome 1.
[2]Relates to Linguistics Department's HLT Program Outcomes 1, 2, & 3.

Instructor

Eric Jackson
Email: `ejackson1@arizona.edu`
Office: (not on campus)
Office hours:

- Open: Thursdays 10am-12pm(`https://arizona.zoom.us/j/87884847069`, pin 808216)

- Appointments: `https://calendly.com/meet_with_eric/60min`

Contact

Students should ask all course-related questions in the course forum (see our D2L page), where you will also find announcements. For emergencies, or personal matters that you don't wish to put in a private post, please email your instructor at `ejackson1@arizona.edu`.

For planning purposes, please note that **your instructor responds to emails and posted questions Monday through Friday from 9AM–5PM MST.** Typically, you can expect a response within a day.

Schedule

This course consists of 6 Units each lasting approximately a week:

Unit	Topic
1	Getting started
2	Tokens and their attributes
3	Tokenization and regular expressions
4	Vector-based representations of words and documents
5	Probability basics
6	Comparing vectors

Assessments Please check D2L for assessment details and due dates:

- `https://d2l.arizona.edu/d2l/home`

Readings   Please check the following URL for the readings schedule:

- `https://d2l.arizona.edu/d2l/home`

Technology

I strongly recommend that you use a computer with at least 8GB of RAM and ≥50GB of free space.

In Unit 1, students will configure a Linux-based development environment using a virtual machine. We will use this development environment for all coursework.

Python-specific Assistance

For students seeking assistance with Python, the University of Arizona Library's Catalyst Studios provides free Python programming drop-in hours throughout the semester. You can find more information and specific times on the library's Catalyst pages:

- `https://libguides.library.arizona.edu/catalyst`

For students entirely new to Python (but not to programming), this free self-paced course can help bring you up to speed quickly:

- `https://arizona.openclass.ai/invite?code=tEgINUEN5ExbEg`

Other Resources

ResBaz Arizona hosts a variety of events (both virtual and in-person) providing researchers opportunities to connect with experienced data scientists and engineers. For more information, see their homepage:

- `https://researchbazaar.arizona.edu/#portfolio`

Readings

The primary text used in this course is freely available (digital-only):

> Dan Jurafsky and James Martin (2019). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* 3rd ed. Upper Saddle River, N.J: Pearson Prentice Hall. ISBN: 9780131873216. URL: `https://web.stanford.edu/%7Ejurafsky/slp3/`

Required readings (papers, chapters, etc.) will be provided by the instructor.

Supplemental Reading

In addition to the course textbook and any posted readings, students may find the following resources useful:

- Michael Hammond (2020). *Python for Linguists*. Cambridge University Press. DOI: 10.1017/9781108642408

- Steven Bird, Ewan Klein, and Edward Loper (2009). *Natural Language Processing with Python*. O'Reilly Media. URL: https://www.nltk.org/book/

Electronic versions of all the recommended resources listed above are freely to University of Arizona students.

2 Evaluation

	Assessment	Description
70%	Programming Assignments	Simple programming exercises designed to deepen your understanding of concepts and techniques covered in each unit (concepts $\rightarrow$ code). Test cases will be provided to help you refine your solutions.
30%	Review & Mastery Assignments	Low-risk assessment consisting of (roughly weekly) guided reviews and questions designed to assist in *retaining and mastering* material covered in each unit.

3 Grading

Grades will be posted to the course's D2L site:

- https://d2l.arizona.edu/d2l/home

For assignments involving code, You will be provided with a subset of test cases to help you refine your solution before submitting.

The grading scheme is as follows:

Grade	Point Range		
A	90	–	100
B	80	–	89
C	70	–	79
D	60	–	69
E	0	–	59

Grade Disputes

Disputes about grades on a particular project will be entertained for two weeks from the day the project is due, or 1 day before grades are due, whichever is sooner. These will be resolved by re-grading the entire project. Note that this can result in a lower grade in the event that new mistakes are discovered.

No negotiations about individual students' letter grades will be entertained once final grades are assigned, except as permitted by the policy stated above.

Collaboration Policy

Students are encouraged to discuss problems and general approaches for solutions. However, **project implementations and the associated documentation for each project must be completed individually** (*except for the final project).* **Copying another person's work (even if it comes from a website) is not permitted and will be treated as a case of academic dishonesty**.

Late Policy

Projects are due electronically via D2L or GitHub Classroom by the stated deadline. No late work will be accepted.

4 University Policies

Missed Classes (Absence)

Accommodation of Religious Observance and Practice:

- `https://policy.arizona.edu/human-resources/religious-accommodation-policy`

All holidays or special events observed by organized religions will be honored for those students who show affiliation with such religions. Absences pre-approved by the UA Dean of Students office will be honored. No matter the reason for missing class, the student is always responsible for the missed material.

With the exception of the above, attendance is mandatory. Students who miss more than $1/3$ **of classes will be dropped. For students in online-only classrooms, attendance means timely completion of assignments and modules as outlined in the course schedule.**

Classroom Behavior

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed.

The Arizona Board of Regents Student Code of Conduct is here:

- `https://public.azregents.edu/Policy%20Manual/5-308-Student%20Code%20of%20Conduct.pdf`

ABOR Policy 5-308, prohibits threats of physical harm to any member of the University community, including to oneself:

- `http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students`

Valuing Inclusion and Diversity

The University of Arizona strives to foster inclusive learning environments in which diversity is recognized and valued. This course strives to create a civil and welcoming environment for everyone, including students of diverse ethnic, cultural, linguistic, national, and familial backgrounds and gender identities, ages, abilities, and veteran status. Students and faculty are responsible for creating an inclusive learning environment through respectful and civil discussion in the classroom and inclusive practices in class and group work.

- We will use the names and pronouns as selected and proposed by each individual in the course in acknowledgment of the intricate nature of our identities and in respect of each other's integrity. Students may request a University wide name change through LGBTQ Affairs at and learn more about name and pronoun use at `https://lgbtq.arizona.edu/transgender-resources`.

- If you have questions about particular policies concerning gender equity, sexual harassment, or sexual assault, please contact the Office of Institutional Equity (https://titleix.arizona.edu) or the Dean of Students Office (`https://deanofstudents.arizona.edu`).

Respectful and Careful Communication

All communication in this class adheres to the principles of civil discourse. Civil discourse is guided by mutual respect and appreciation. Diversity of knowledge is an asset to class discussions. In all communication, you are expected to be scholarly, professional, and respectful. Constructive criticism in discussion of course concepts is highly encouraged. Mocking and/or bullying are never allowed. To be critical does not exclude being polite. See:

- UA National Institute of Civil Discourse: `https://nicd.arizona.edu`

- UA Dean of Student's page on cyberbullying: `https://deanofstudents.arizona.edu/safety/cyberbullying`

Accessibility and Accommodations

At the University of Arizona, we strive to make learning experiences as accessible as possible. If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, `https://drc.arizona.edu`) to establish reasonable accommodations.

Student Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described here:

- `https://deanofstudents.arizona.edu/policies/code-academic-integrity`

Confidentiality of Student Records

`https://registrar.arizona.edu/personal-information/student-information`

On Dropping Classes

If you find yourself thinking about dropping this (or any other) class, first make sure that that's what you really want to do. Chatting with the instructor or your academic advisor may help. If you drop within the first week of the term, there will be no notation on your transcript; it will be as though you'd never enrolled. After the second week, a drop will be recorded on your transcript. You will receive a "WP" (withdrawn passing) only if you were passing the class at the time of your drop. Toward the end of the term, dropping becomes a challenge, because you need to explain to the instructor and to the dean why you were unable to drop the class during the first half of the term. For drop deadlines specific to this compressed format, please see this calendar:

- `https://www.registrar.arizona.edu/dates-and-deadlines`

Subject to Change Statement

The instructor reserves the right to change with advance notice where appropriate the content of the course. This right does not apply to posted grading and absence policies or University Policies.