LING 531
Jackson
Fall 2022

Human Language Technology II

Course information

This intermediate-level course is a continuation of LING 529 and covers the basics of information retrieval, focusing on both search and classification.

Course objectives

This course will present students with the fundamentals of text search in the context of a simple boolean search. We'll then refine our methods for effective search—with the goal of returning the *best* results—by exploring issues of similarity and weighting of terms. We'll finish the course by exploring document classification, comparing statistical methods and vectorspace methods.

Learning outcomes

Successful students in this course will...

1. use Python and tools commonly used in commercial applications to perform simple but realistic information retrieval tasks.
2. parse and prepare a corpus for search and classification tasks.
3. create an index for searching a corpus.
4. implement similarity and evaluation measures for searching a corpus.
5. implement classification models using statistical and vectorspace methods.
6. compare and evaluate these search and classification methods and the results they yield.

Learning outcomes #1 and #2 relate to Linguistics HLT program outcomes #1 and #3. Learning outcomes #3-6 relate to Linguistics HLT program outcome #2.

HLT learning outcomes addressed in this course

1. Students will demonstrate programming skills for the workplace.
2. Students will be able to use fundamental algorithms and concepts in Natural Language Processing.
3. Students will show knowledge of tools and packages used in Natural Language Processing.

Prerequisites

Ling 529 or equivalent.

Instructor

name	Eric Jackson
email	ejackson1@email.arizona.edu
office hours	Thurs. 10:00am–12:00pm (Arizona time, UTC-7) and by appointment, online via Zoom at https://arizona.zoom.us/j/88095722750 (passcode 975869)
appointments	`https://calendly.com/meet_with_eric/60min`

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 I respond to emails and questions posted on the forum **Monday through Friday from 9am to 5pm MST**. I am not available for course-related interaction on weekends. Typically, you can expect a response from me within one working day.

Topics and schedule

The schedule below should be interpreted as follows:

- For readings, you should complete the indicated reading *before* the week begins.
- Assignments are all due Tuesday at noon Arizona time for the week indicated. (Exceptionally, the final assignment is due on a Thursday.)
- Each topic has associated videos and Jupyter Notebooks; you should go through these for the week indicated.

The first week, last week, and Thanksgiving week (11/24 & 11/25 are holidays) are all treated as **half weeks** where there are reduced expectations for workload. An asterisk * indicates these short weeks and short assignments, "R&M" is a *Review & Mastery* activity, "Disc" is a *forum discussion post*, and "HW" is a programming assignment, to be explained below.

Week	Dates	Topic	Reading	Notebook	Due
1	10/14-10/15	Overview*	IR ch.1	#1	R&M 1, Disc 1
2	10/18-10/22	Indexing	IR ch.2	#2,3	HW 1*, R&M 2, Disc 2
3	10/25-10/29	Similarity	IR ch.3	#4	HW 2, R&M 3, Disc 3
4	11/1-11/5	Weighting	IR chs.4,6	#5	HW 3, R&M 4, Disc 4
5	11/8-11/12	Measuring	IR chs.7,8	#6,7	HW 4, R&M 5, Disc 5
6	11/15-11/19	Classifying	IR ch.13	#8	HW 5, R&M 6, Disc 6
7	11/22-11/26	Naive Bayes*	IR ch.14	#9	HW 6, R&M 7, Disc 7
8	11/29-12/3	Rocchio/kNN	none	#10	HW 7*, R&M 8, Disc 8
9	12/6-12/8	Catch-up*	none	#11	HW 8, Disc 9

Readings

Draft versions of the course textbooks are available for free online. Assigned readings come from the first; the second is recommended for reference and should be familiar from HLT 1.

- *Introduction to Information Retrieval*, Manning, Raghavan, and Schütze, `https://nlp.stanford.edu/IR-book/information-retrieval-book.html`
- *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Jurafsky and Martin, `https://web.stanford.edu/%7Ejurafsky/slp3/`

Required assignments and course grading

	Assessment type	**Description**
60%	Programming assignments *(6 regular @ 8% each,* *2 small @ 6% each)*	Simple programming exercises in Jupyter Notebooks designed to deepen your understanding of concepts and techniques covered in each unit. Test cases will be provided to help you refine your solutions.
30%	Review & Mastery activities *(8 assignments, graded only for completion)*	Low-risk assessment consisting of a guided review and questions each week. These are designed to help you **retain and master** material covered in each unit.
10%	Forum interaction *(9 posts graded for completion)*	You will be given an opportunity to choose a topic to write on, with no length requirement, based on the week's reading and topics. You are encouraged to read and respond to other students' posts, as well.

Programming assignments will be due every Tuesday at noon, Arizona time, except for #8 which will be due on a Thursday. Submission of your programming assignment will involve submitting your Jupyter Notebook file in the appropriate assignment item in D2L.

Review & Mastery activities and forum posts will be due by Monday morning at 9am, Arizona time. For Review & Mastery activities, your progress and times are tracked in the OpenClass.ai system and synchronized in D2L after the activity due date.

Forum posts will be made on the course forum (not in D2L) in a distinct channel for each unit, at `https://forum.hlt.arizona.edu/`. I will manually check for posts and update forum post credit in D2L during my normal working hours.

Authoritative due dates, and grades for each item, will be posted in D2L.

If you are struggling with a topic or an assignment, please contact me for help well before the due date. I want to help you understand and master this content, but I can only do that if you communicate with me about problems that you encounter. Since this is a 3-credit course compressed into half a semester, it's important to quickly address problems of understanding and keep up with assignments. **Late work will otherwise not be accepted.**

Technology

You'll need to access to a linux-like environment including Python 3 and Jupyter. Those of you continuing from HLT 1 are free to use your linux virtual machine from that course.

If you prefer, you may install Python 3 and Jupyter in your host OS. One way to accomplish this is to use Anaconda, a Python distribution that is free and available for all platforms. You're welcome to use another Python distribution besides Anaconda if you prefer, including one that is preinstalled in Ubuntu or macOS, but it's your responsibility to make sure it does everything that our class notebooks require.

`https://www.anaconda.com/products/individual`

I expect you to have access to Python 3 and Jupyter Notebooks either directly in your system (whether Windows or macOS) or indirectly via Docker, Virtualbox, a dual-boot Linux system, etc. If you have problems with this, please talk to me as soon as possible.

Collaboration Policy

Students are encouraged to discuss problems and general approaches for solutions, but everyone must turn in their own work. You may not submit assignments that are substantially the same as your classmates.

University boilerplate

All of the following items are required by the university to be included on syllabi. If you find something here that is surprising or unexpected, please bring it up with me as soon as possible.

By way of a brief summary:

Disabilities If you have a disability that affects how you will need to do the work in this class, please let me know *within the first week of class.*

Academic Code of Conduct Cheating and plagiarism are not remotely acceptable in any way. Disruptive behavior in class—which here means on any of our course websites or by email —is not acceptable. Please be respectful of others.

Sensitive Material This is a university and you are adults. It is possible that we may touch on topics that some students could find sensitive during the semester. Given the focus of this course, this seems unlikely, but I alert you nonetheless.

Covid

The university has a specific site for covid information: `http://covid19.arizona.edu`. These are extraordinary times and you may still be experiencing personal and financial challenges. Let me know if we need to make accommodations for covid-related things, and please stay safe.

Absence and Class Participation Policy

Attendance in an all-online course is not evaluated like attendance in an in-person course. For this course, attendance will be represented by active reading, completion, and participation in online course activities, including materials and activities posted on D2L, OpenClass, our course forum, and any other related websites.

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at: http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop

The UA policy regarding absences is that any sincerely held religious belief, observance or practice will be accommodated where reasonable, http://policy.arizona.edu/human-resources/religious-accommodation-policy.

Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: https://deanofstudents.arizona.edu/absences

Classroom Behavior Policy

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. To that end, our focus is on the tasks at hand and not on extraneous activities.

Students are asked to refrain from disruptive conversations with others in the course. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students.

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.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. If you found a code snippet online, it's important to cite where it came from, even if that source was stackexchange.com.

Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: `http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity`.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see

`http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy`

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.