Ling 531
Jackson
Fall 2021

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—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 #3-5 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
hours	Tues. 7:00pm–9:00pm (Arizona time, UTC-7) and by appointment, online via Zoom at `https://arizona.zoom.us/j/83707417647` (passcode 150782)

Requirements

There will be six regular weekly homework assignments (#2, #3, #4, #5, #6, #8) and two smaller weekly assignments (#1 and #7), worth the values below:

type	number	value each	total
regular assignments	6	14%	84%
smaller assignments	2	8%	16%
total			100%

These assignments will be due every Tuesday at noon (Arizona time) except for #8 which will be due on a Thursday.

Late work will otherwise not be accepted.

Readings

Draft versions of the textbooks used in this course are available for free on-line. The first below is required and the second recommended.

- *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/`

Technology

You'll need to access to a linux-like environment including python 3 and jupyter (ideally Anaconda). Anaconda is free for all platforms and I'll assume you have that functionality either directly in your system or indirectly via docker, virtualbox, a dual-boot linux system, etc. You're welcome to use another python distribution if you prefer, but it's your responsibility to make sure it does everything that the anaconda one does.

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

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.

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.
- 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 (11/25 & 11/26 are holidays) should all be treated as *half weeks** where there will be reduced expectations as far as workload.

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

Covid

1. The university has a specific site for covid information: `http://covid19.arizona.edu`.
2. I understand that these are extraordinary times and folks are experiencing new personal and financial challenges. Let me know if we need to make accommodations for covid-related things.
3. . . . and please stay safe.

Boilerplate

Disabilities If you have a disability that affects how you will need to do the work in this class, please let us 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 is not acceptable.

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.

University boilerplate

All of the following are things the university requires us to put on syllabi.

Absence and Class Participation Policy

Attendance is not required, but the university requires the following on syllabi anyway.

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 for 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 (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Students are asked to refrain from disruptive conversations with people sitting around them during lecture. 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

Required language: 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

Recommended language is provided on the Disability Resource Center website: http://
drc.arizona.edu/instructors/syllabus-statement.

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. 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.