LING 3802 - Language and Computers Syllabus - AU2018

TR 9:35-10:55am
Baker Systems 285

Instructor: Evan Jaffe
Email: jaffe.59@osu.edu
Office Hours: Oxley Hall 200B, W 1:30-2:30pm, R 2:30-3:30pm, and by appointment

If you have any concerns about myself or about the course that you prefer to not discuss with me directly, feel free to contact the Linguistics teaching coordinator with the information below:
GTA Coordinator: Dr. Hope Dawson
Email: dawson.165@osu.edu
Office: 114 Oxley Hall
Phone: 292-5420

Course Description:
For many people, increasing computer use in personal and professional spheres has revolutionized the way we communicate, play, and work. Natural language is a convenient medium for exchanging information, so text and speech representations (encoding formats) are critical in getting computers to process and organize information.

This course introduces fundamentals of human language technology, including web search, spam filtering, spelling and grammar checkers, dialogue systems, automatic machine translation, voice recognition and more - discussing both how they work *and* why they often don't. To a lesser extent, we will also examine social and ethical issues of privacy, misinformation, job creation/loss due to language technologies and the nature of machine intelligence.

GE fulfillment:
This course satisfies the GE category Quantitative Reasoning, Mathematical and Logical Analysis. The goals of this category are for students to develop skills in quantitative literacy and logical reasoning, including the ability to identify valid arguments, and use mathematical models. The expected learning outcomes are for students to comprehend mathematical concepts and methods adequate to construct valid arguments, understand inductive and deductive reasoning, and increase their general problem solving skills.

This course satisfies the goals and learning outcomes by using natural language systems to motivate students to exercise and develop a range of basic skills in formal and computational analysis. The course philosophy is to ground abstract concepts in real world examples. We introduce strings, regular expressions, finite-state and context-free grammars, as well as probabilistic algorithms defined over these structures and techniques for probing and evaluating systems that rely on these algorithms. The course goes beyond merely subjective evaluation of

systems, emphasizing quantitative analysis and reasoning to draw are argue for valid conclusions about the design, capabilities and behavior of natural language systems.

Carmen:

Carmen will be used for the schedule, homework and reading assignments. Check Carmen regularly to stay up to date on assignments. You can enable email notifications for Carmen. There will also be discussion forums for posting questions and providing course feedback (anonymously, if desired).

Readings:

Language and Computers. by Markus Dickinson, Chris Brew, Detmar Meurers
ISBN: 978-1-405-18305-5. Oct 2012, Wiley-Blackwell.
Textbook readings are mandatory and include quizzes before the material is presented in class. During class, material will be presented that relies on the basics from the reading, so come prepared. Class materials will be posted on Carmen, as well as slides, but these are not a replacement to being in class! We will also be using the Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit by Steven Bird, Ewan Klein, and Edward Loper, which is available online. Other readings may also be assigned and posted on Carmen.

Requirements:
- Quizzes (5%):
  Quizzes will be administered online through Carmen and are due before class (<9:35am) of the day indicated, to be completed individually - no collaboration. They will open after the class meeting preceding their due date. E.g., a Tuesday quiz will open the previous Thursday. The quizzes are open book, but you should finish the reading before attempting them as only one attempt is allowed and they are timed (15 mins). They will be shut off automatically once the deadline is reached, so do not put it off to the last minute! Note that I do not promise to remind you when you have a quiz due; it is your responsibility to keep up with the schedule on Carmen. Answers will appear after the quiz has closed.

- Homework assignments (35%):
  Homework assignments are due by the beginning of class via Carmen. PDF format is preferred. Please do not email me your homework; it will not be accepted. No late homeworks will be accepted unless you make arrangements with me ahead of time. Homework problems are typically similar to ones explored in groups during class; as such, regular attendance and active participation is vital to doing well on the homeworks. Homeworks can be discussed with other students, however the write-up and final product must be done **individually** and you **must** include the names of the people you discussed/collaborated with. I am required to report any suspected violation of academic conduct - see Academic Misconduct Policy below. Homeworks are due by midnight of

the day on the calendar.

- Essay (15%):
  A 1000–1500 word essay on a topic dealing with the social implications of language technology. This will be due after Thanksgiving - if you don't want to spend your break writing the essay, I encourage you to work ahead and plan accordingly. A detailed description and rubric will be posted on Carmen.

- Midterm exam (20%):
  The midterm will be given in class on Tuesday, October 9 and covers the first half of course material. Closed book.

- Final exam (20%):
  The final will be given on Friday, December 7 (8:00am–9:45am). The final is cumulative but will focus on the second half of the course material. Closed book.

- Class participation (5%):
  Given that the homeworks and exams reflect the material covered in class, attendance is essential for doing well in this course, as is your active participation in class discussion and in-class activities.

- Grades will be assigned using the standard OSU scale: 93-100 (A), 90-92.9 (A-), 87-89.9 (B+), 83-86.9 (B), 80-82.9 (B-), 77-79.9 (C+), 73-76.9 (C), 70-72.9 (C-), 67-69.9 (D+), 60-66.9 (D), below 60 (E). For hundreths of a point, .05 and above will be rounded up, .04 and below will be rounded down.

Schedule:

Week	Month	Date	Day	Topic	Assignments Due
1	Aug	21	T	Introduction	
		23	R	Speech Encoding (acoustics, n-grams)	Chapter 1
2		28	T	Text Encoding (alphabets, ascii, numeral systems)	Quiz 1
		30	R		HW 1
3	Sep	4	T	Searching (web search, evaluation, FSAs, Regexp, corpora stats, Jeopardy!)	Chapter 4 Quiz 2
		6	R		

4		11	T		HW 2
		13	R	Text Classification (phishing, fake news, spam filtering, sentiment analysis, paraphrase detection, Bag-of-words, tf-idf, classification algorithms)	Chapter 5 Quiz 3
5		18	T		
		20	R		HW 3
6		25	T	Writer's Aids (Spell-checking, edit distance, syntactic context, N-gram models, ontologies, Grammarly)	Chapter 2 Quiz 4
		27	R		
7	Oct	2	T		
		4	R	Review	HW 4
8		9	T		**MIDTERM**
		11	R	FALL BREAK	
9		16	T	Machine Translation (Rosetta Stone, IBM Model, current systems)	Chapter 7 Quiz 5
		18	R		
10		23	T		HW 5
		25	R	Dialog Systems (discourse management, ELIZA, Siri/Alexa/Cortana etc.)	Chapter 6 Quiz 6
11		30	T		
	Nov	1	R		HW 6
12		6	T	Computer-aided language learning (acquisition, tutoring systems, Duolingo, Rosetta Stone, Rikai-kun, Spritz, etc.)	Chapter 3 Quiz 7
		8	R		
13		13	T	Social Context of Technology Use (Privacy, Consciousness and	Chapter 8 Quiz 8

				Intelligence, Job Loss/Creation, etc.)	
		15	R		HW 7
14		20	T	Guest lecture (schedule permitting)	
		22	R	THANKSGIVING	
15		27	T	Advanced Topics (Word Embeddings, Current Research)	Essay Due
		29	R		
16	Dec	4	T	Review	
		7	F	8am-9:45am (NOT regular class time)	**FINAL**

Exams and Important Dates:
MIDTERM will be on Tuesday, October 9 in class and covers the first half of course material.
FINAL will be on Friday, December 7, 8am-9:45am and is cumulative, but focused on second half of course material.

Make-up Policy:
If you know you will miss a deadline or exam, please talk to me **before** you miss it, and we can try to accommodate the absence. If you miss the midterm or final, you will need to provide extensive written documentation for your excuse.

Etiquette:
Students are expected to be respectful at all times during class. Disagreement is fine, but it should voiced in terms of the material, not as personal attacks. Treat others as you'd like to be treated. Additionally, while in class please do not do non-course-related activities, including texting, talking on the phone, other coursework, watching cat videos, etc. Unless we're discussing language technology as it relates to cat videos.

Academic Misconduct Policy:
As with any class at this university, students are required to follow the Ohio State Code of Student Conduct (https://trustees.osu.edu/assets/files/RuleBook/CodeStudentConduct.pdf). In particular, note that students are not allowed to, among other things, submit plagiarized (copied but unacknowledged) work for credit. I am required to report any suspected violation to the Council on Academic Misconduct. See the Office of Academic Affairs information for students (https://oaa.osu.edu/academic-integrity-and-misconduct/student-misconduct).

Students with Disabilities:
Students who need an accommodation based on the impact of a disability should contact me to arrange an appointment as soon as possible to discuss the course format, to anticipate needs,

and to explore potential accommodations. I rely on the Student Life Disability Services office for assistance in verifying the need for accommodations and developing accommodation strategies. Students who have not previously contacted Student Life Disability Services are encouraged to do so (614-292-3307, http://slds.osu.edu/)

Counseling and Consultation:
If you are having any personal issues and would like to talk to someone, there are resources available for you. You can contact Student Life's Counseling and Consultation Service at 614-292-5766
- Counseling is confidential
- Counseling does not affect academic records
- Counseling sessions are free to registered Ohio State students

Disclaimer:
This syllabus is subject to change. All important changes will be announced in class, on Carmen, and by email.