CS 322.00 – Natural Language Processing

Fall 2017

3a (M, W 11:10-12:20pm: F 12:00-1:00pm)

Leighton 305

Instructor:	Blake Howald
Office:	Center for Math and Computing (CMC) 134
Email:	bhowald @ carleton.edu
Phone:	x6194
Office Hours:	Monday 3-5p; Thursday 1:30-3:30p; by appointment

Grader:	Camden Sikes
Email:	sikesc @ carleton.edu

Course Description:

The ability to acquire and communicate with a natural language is one of the things that makes humans unique as a species. However, getting computers to do language the way people do language, which is a key area of focus in artificial intelligence, is a really hard problem. This course will provide an overview of the linguistic theories and computational techniques developed since the mid-20th century that have made many aspects of the interaction between computers and natural language very much a reality – collectively referred to as Natural Language Processing or NLP.

Topics in this course include morphological and syntactic processing, semantic analysis, document classification, speech recognition, dialogue systems, machine translation and more. Particular emphasis will be placed on (1) understanding the underlying linguistic and computational motivations for different techniques; (2) practical applications and implementations of core algorithms and methods; (3) analysis of data; and (4) evaluation of system (and human) performance.

CS 201 – Data Structures is a required prerequisite. Additionally, either CS 202 – Mathematics of Computer Science, Math 236 – Mathematical Structures, or instructor permission.

The programming language of choice for this course will be Python (https://www.python.org) - Please visit this link to install on your personal computer if you so choose (https://apps.carleton.edu/curricular/cs/resources/source/python_install/). Otherwise, everything you may need (including libraries such as NLTK) will be installed on the lab machines. Contact Mike Tie (mtie@carleton.edu, x4067, CMC305) if you require assistance.

REQUIRED TEXT AND READINGS:

Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd Edition). Daniel Jurafsky and James Martin. Pearson/Prentice Hall ("**J&M**").

Additional readings, resources and discussions will be posted on Moodle.

GRADING* AND REQUIREMENTS:

Programming Assignments [4]		40% (4X10%)
Quizzes		10% (2x5%)
Midterm		10%
Final		10%
Group Research Project:		20%
Proposal	2.5%	
Progress Report	2.5%	
Presentation	5%	
Final Write-Up	10%	
Participation (Reading Presentation [5%])		10%
		100%

PROGRAMMING ASSIGNMENTS:

There will be four programming assignments (each worth 10% of your overall grade). Each assignment is designed to reinforce algorithmic and implementation details of the theoretical concepts covered in class and will be multi-dimensional in nature – in particular, each assignment will require the processing of textual data, coding of methods, computation of results, error analysis and response.

Collaboration is strongly encouraged and you can work with ***one*** person on any of the homework assignments. If you do work in pairs, only one of you should submit the assignment indicating both contributors. You will have roughly a week and a half to complete the assignments, please start early.

Code and any supporting write-ups and data will be submitted by 11:59 via Moodle on the day that it is due. Programming assignments submitted after the deadline, but within 24 hours, will be docked 50%. After that, you're out of luck. If you know you're going to have trouble with an approaching deadline, please let me know ASAP and we'll work something out.

* Our grader Camden Sikes is also enrolled in the class. To avoid conflicts of interest and bias, any work graded by Camden will be anonymized by me in advance of grading (I will be grading Camden's work). Please avoid identifying yourself (or yourselves, if collaborating) in the assignment. More details will be given in class.

MIDTERM AND FINAL EXAMS:

There will be two relatively short in-class quizzes (5% each, Weeks 3 and 7) bridging the gaps between the beginning of class and the midterm exam and the midterm and final exams.

There will be an in-class Midterm given on ***Friday, October 13, 2017,*** and an in class Final given on ***Wednesday, November 15, 2017.*** You will have the entire class period (and only the class period) to complete the exams.

All quizzes and exams will be comprehensive (covering all material prior to the date of the quiz or exam) and will assess your understanding of the core theoretical concepts and associated implementation considerations. The quizzes and exams will consist of a combination of short answer/essay questions and data/algorithm/system/results analysis.

GROUP RESEARCH PROJECT:

Throughout this course you will work in groups of no more than 4 total class members to define, plan and execute a research project. While the topics can range considerably, the overarching format will remain the same for all groups: (1) Define a problem; (2) Identify data to be processed in furtherance of addressing the problem; (3) design and program a solution to the problem that analyzes or transforms the data by leveraging NLP concepts and tools; and (4) provide a final write-up of the project to include a one page ethical statement on the impacts of your group's research.

Within the first two weeks of the term, groups should be defined and initial rough proposals (2.5%) submitted for review (Week 2). Once approved, work can begin and a progress report (2.5%) will be due shortly after the Midterm break (Week 6). 20 minute presentations (5%) will then be scheduled and given in Week 9. By 5 PM **Monday, November 20, 2017**, the group will jointly submit a final report of the project, data, code, ethics impact statement and a group evaluation report.

PARTICIPATION:

Active participation in this course, which can take the form of in-class and on-line discussions, small group activities and completing readings among other things, is essential. Half of the Participation grade allocation (5%) will be based on a joint (2 person) **15 minute in-class presentation** (10 minutes for presentation, 5 minutes for questions) on a topic of your choosing – possibilities include (1) NLP tools (open source or otherwise); (2) datasets; (3) an article from the ACL Anthology (http://aclanthology.info); (4) a key work referenced in our reading. Please post any readings or materials associated with your presentation 24 hours before you are scheduled to present. Let me know if you need some guidance. **I will circulate a sign-up schedule in Week 1.** Because of the size of the class, the first presentation will be **Wednesday, September 20, 2017**.

COURSE OUTLINE: Please complete readings in advance of the lecture date so as to maximize the benefit of lecture and discussion.

Week	Date	Topic	Reading	Important Dates
1	9/11	Introductions	J&M Chapters 1 & 2	
	9/13	Automata		
	9/15	Regex / IR Evaluation		HW 1 Assigned (9/15)
2	9/18	Morphology	J&M Chapter 3	
	9/20	Morphology (con't)		
	9/22	N-Grams	J&M Chapter 4 thru 4.7 (skip 4.5.3 & 4.7.1)	Project Ideas Due (9/22)
3	9/25	N-Grams (con't)		HW 1 Due (9/25)
	9/27			
	9/29	Quiz 1, Evaluations		Quiz #1/HW 2 Assigned (9/29)
4	10/2	POS Tagging	J&M Chapter 5 thru 5.7.1	
	10/4			
	10/6	Grammars	J&M Chapter 12	
5	10/9	Machine Translation, Ethics	Russell and Norvig (2010) §§26.3-.4	HW 2 Due (10/9)
	10/11	MT(con't)/Review	J&M Chapter 25 thru 25.9 (skip 25.6.1 thru 25.8)	HW 3 Assigned (10/11)
	10/13	**Midterm Exam**		
6	10/16	**Midterm Break**		
	10/18	Syntactic Parsing	J&M Chapters 13	
	10/20	Syntactic Parsing (con't)	J&M Chapter 14 thru 14.3, 14.7, 14.10)	Progress Reports Due (10/20)
7	10/23	Semantics	J&M Chapter 17	HW 3 Due (10/23)
	10/25	Semantics (con't)	J&M Chapter 19	HW4 Assigned (10/25)
	10/27	Pragmatics/ Discourse	J&M Chapter 21	Quiz #2
8	10/30	Annotation	Palmer & Xue (2013)	
	11/1	Speech Processing	Renals & Hain (2013)	
	11/3	Speech Processing (con't)		HW 4 Due (11/3)
9	11/6	Presentations (All Week)		
	11/8			
	11/10			
10	11/13	Catch-Up/Review		
	11/15	**Final Exam**		

THE WRITING CENTER:

The Writing Center, located in 420 4th Libe, has peer writing consultants who can work with you during any stage of the writing process (brainstorming to final proofreading). Hours and more information can be found on the writing center website. You can reserve specific times for conferences in 420 4th Libe by using their online appointment system. Walk-ins are welcome, though writers with appointments have priority.

If you are a second language writer and believe you might benefit from working individually with a writing consultant on a regular basis this term, email Renata Fitzpatrick, Multilingual Writing Coordinator, call her at x5998, or stop by her office in 420D 4th Libe. She can arrange once- or twice-a-week meetings between you and a specific writing consultant throughout the term.

CLASS PRESENTATIONS AND PUBLIC SPEAKING SUPPORT:

Speech coaching is a student-staffed resource designed to assist you with class presentations, comps talks, and other speech-related events. Your coach can assist you with speech & communication skills including clarity, organization, articulation, projection, body language, eye contact, and effective use of aids (e.g., notes, PowerPoint, Keynote, etc.). Depending on your goals, your coach can also work with you on the content of the presentation: organization, voice, clarity, and, ultimately, persuasive impact. Individuals and groups are welcome to request a speech coach by completing a brief, online form. The speech coach will meet you at a mutually convenient time and place. For more information, visit go.carleton.edu/speakeasy.

TIME MANAGEMENT, TEST PREPARATION STRATEGIES, AND STUDY SKILLS:

All Residential Life Area Directors are trained to work with you to improve your time management and academic skills. Their goals are to heighten your awareness of your personal strengths and skills and to offer different ways you can approach your academic work so you're more efficient and effective. Meetings are by appointment; you simply need to email one of them to arrange a visit. For details and resources, see the Learning Strategies Coaching website.

THE LIBE:

Ask a librarian for help with your research in this class. The library liaison for Linguistics is Iris Jastram. You can drop by the library's Research/IT desk to ask any question you have, at any point in your process. Librarians help students find and evaluate articles, books, websites, statistics, data, government documents, and more. For more information on hours and librarians, visit the Gould Library website at go.carleton.edu/library.

ACADEMIC HONESTY[1] (EXCERPTED FROM THE 'ACADEMIC INTEGRITY' PAGE OF CARLETON'S ACADEMIC REGULATIONS AND PROCEDURES HANDBOOK):
It is assumed that a student is the author of all course work (quizzes, problem sets, online contributions, tests, papers, lab work, etc.) that he/she submits, whether for a grade or not, and that the work has not been submitted for credit in another class without the instructor's permission. Images, ideas, data, audio clips, or phrases borrowed from others should be fully identified by standard procedures for making such acknowledgment. All permitted collaboration with others must still be acknowledged… At Carleton College, an act of academic dishonesty is therefore regarded as conflicting with the work and purpose of the entire College and not merely as a private matter between the student and an instructor; all cases involving such dishonesty are referred for appropriate action to the Academic Standing Committee (ASC) via the Associate Dean of Students or the Associate Dean of the College.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES:
Carleton College is committed to providing equitable access to learning opportunities for all students. The Disability Services office (Burton Hall 03) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations. If you have, or think you may have, a disability (e.g., mental health, attentional, learning, autism spectrum disorders, chronic health, traumatic brain injury and concussions, sensory, or physical), please contact Chris Dallager, Director of Disability Services, by calling 507-222-5250 or sending an email to cdallager@carleton.edu to arrange a confidential discussion regarding equitable access and reasonable accommodations.

ASSISTIVE TECHNOLOGY: The Assistive Technology program brings together academic and technological resources to complement student classroom and computing needs, particularly in support of students with physical or learning disabilities. Accessibility features include text-to-speech (Kurzweil), speech-to-text (Dragon) software, and audio recording Smartpens. If you would like to know more, contact aztechs@carleton.edu or visit go.carleton.edu/aztech.

[1] For more information on Carleton's policies surrounding academic honesty, please refer to the *Academic Integrity in the Writing of Essays and Other Papers* (http://apps.carleton.edu/campus/doc/integrity/). As always, feel free to speak with me about any questions or concerns you may have.