W&M DATA 340 4 Natural Language Processing, Fall 2023 Syllabus

Schedule | Email Instructor

Course overview

This course is designed to introduce students to Natural Language Processing (NLP) and its applications in academic research, data science, and industry. Students will learn how to use natural language processing techniques to gain a deeper understanding of a research question and/or topic.

Course venue and time

- Integrated Science Center, room 1280. Tuesdays (T) and Thursdays (Th) 5:00 - 6:20
- See course in Blackboard

Instructor contact and office hours

- email: jmtucker02@wm.edu or j.tucker@mail.utoronto.ca
- website: https://jamesmtucker.com
- Office hours by appointment only

Programming language

This course is language agnostic. You can submit your homework and project in whatever programming language you prefer. In class lectures, we will use Python, R, or Mojo for the most part.

Course objectives

- Understand the basics of natural language processing techniques and how they can be used to build programs that model human use of language and generate human language
- Learn how to use natural language processing tools and libraries to perform tasks such as text classification, sentiment analysis, and text generation
- Develop the ability and experience to design and implement natural language processing systems for real-world applications
- Explore ethical and social implications of natural language processing and artificial intelligence

Course topics

1. Introduction to natural language processing
2. Data set creation and documentation
3. Text preprocessing and cleaning
4. Text classification and sentiment analysis
5. Neural Networks, Transformers, Large Language Models
6. Ethical and social implications of natural language processing and artificial intelligence

Textbook

Required

- Jurafsky, Dan and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Online: https://web.stanford.edu/~jurafsky/slp3/. PDF

Recommended

- Arcila Calderon, Carlos, et al. *Computational Analysis of Communication.* United Kingdom, Wiley, 2022. Google Books
- Tunstall, Lewis, Leandro von Werra, and Thomas Wolf. *Natural Language Processing with Transformers* O'Reilly Media, 2022. Google Books

Assignments

All assignments are posted cross posted on the Discord channel and GitHub Repo.

- NLP Problem sets (35%) - these are smaller assignments designed to reinforce the ideas discussed in lecture or the assigned reading. These assignments will be posted on the course GitHub repo and Blackboard and are due at the specified date and time.
- Project (50%) - The student will develop a project related to their area of interest and use NLP techniques to explore a data set in consideration with some research questions.
  - Project code, documentation, and report (40% of the 50) - in the overall development of the project. It is hoped that the student will create a GitHub repo to host their code and documentation. The student should expect to submit project code that is well-documented and reproducible. The research findings should be presented in a report that is well-written and well-organized.
  - Project presentation (10% of the 50) - In the final class sessions, each student will present a short presentation of their projects.
- Course preparation (15%) - It is expected that the student will come to class with having read the assigned readings and/or other additional code documentation.

Schedule

date	day	topic	reading	academic calendar
2023-08-31	Th.	Introduction & Syllabus		
2023-09-05	Tu.	NLP, Data Science, & Large Language Models		
2023-09-07	Th.	Stochastic Parrots & Human Communication		
2023-09-12	Tu.	Statistics & Information Theory		Withdrawal period begins
2023-09-14	Th.	Linguistic structure (Pt. I)	[JM 17, 18]	
2023-09-19	Tu.	Linguistic structure (Pt. II)	[JM 17, 18]	
2023-09-21	Th.	Datasets and Data Munging (Pt. I)	[G] & [JM 2]	
2023-09-26	Tu.	Datasets and Data Munging (Pt. II)	[vanA]	
2023-09-28	Th.	Vector semantics (Pt. I)	[JM 6]	
2023-10-03	Tu.	NLP & Machine learning (Pt. I)	[JM 4, 5]	
2023-10-05	Th.	NLP & Machine learning (Pt. II)	[JM 4, 5]	
2023-10-10	Tu.	NLP & Machine learning (Pt. III)	[JM 4, 5]	Midterm grading period
2023-10-12	Th.	Vector semantics (Pt. II)	[JM 6, 7]	Midterm grading period

date	day	topic	reading	academic calendar
2023-10-17	Tu.	NLP & Neural nets (Pt. I)	[JM 7]	Midterm grading period
2023-10-19	Th.	NLP & Neural nets (Pt. II)	[JM 7, 9]	Midterm grading period
2023-10-24	Tu.	NLP & Neural nets (Pt. III)	[JM 10]	Midterm grading period
2023-10-26	Th.	NLP & Neural nets (Pt. IV)	[JM 11]	Midterm grading period
2023-10-31	Tu.	NLP & Neural nets (Pt. V)	[JM 11]	Advising period
2023-11-02	Th.	Vector semantics (Pt. III)		Advising period
2023-11-07	Tu.	No class		Election Day
2023-11-09	Th.	Machine translation	[JM 13]	
2023-11-14	Tu.	QA & Information retrieval	[JM 8, 14]	Spring 2024 priority registration
2023-11-16	Th.	Named Entity Recognition (NER) & Ontologies		Spring 2024 priority registration
2023-11-21	Tu.	Vector semantics (Pt. IV)		Remote instruction days
2023-11-23	Th.	No class		Thanksgiving Break
2023-11-28	Tu.	Vector semantics & Clustering		
2023-11-30	Th.	Student presentations		
2023-12-05	Tu.	Student presentations		
2023-12-07	Th.	Student presentations		
2023-12-12	Tu.	Student presentations*		Final exam periods
2023-12-14	Th.	Student presentations*		Final exam periods
2023-12-19	Tu.	Student presentations*		Final exam periods

* We will only meet on the once during the exam week to present final projects, if necessary.

Course policies

Please read and take notice of the following:

Grade scale

	Mark		Mark
93 - 100	A	73 - 76	C
90 - 92	A-	70 - 72	C-
87 - 89	B+	67 - 69	D+

	Mark		Mark
83 - 86	B	63 - 66	D
80 - 82	B-	60 - 62	D-
77 - 79	C+	00 - 59	F

Grading appeals

To appeal a grade, schedule a meeting to discuss it with me.

Communications

The discord channel provides an additional way to collaborate with our Natural Language Processing community. You are highly encouraged to participate in discussions on the discord channel. If you ask for help, please provide code examples so that the community can better help and understand.

In addition to discord, we will use Blackboard to post grades and announcements. Please check Blackboard regularly for announcements.

The course readings, data sets, and code are available on the course GitHub repo.

Absences

If you are absent please email me and let me know or send me a text message. Course work is due as detailed in the course schedule. Late work is penalized 2% of the earned mark for every day it is late. If you are absent on a day that an assignment or project milestone is due, please make sure to turn it in early. If you are ill, please communicate with me regarding an extension.

Mental Well-Being

William & Mary recognizes that students juggle different responsibilities and can face challenges that make learning difficult. There are many resources available at W&M to help students navigate emotional/psychological, physical/medical, material/accessibility concerns, including:

- The W&M Counseling Center at (757) 221-3620. Services are free and confidential.
- The W&M Health Center at (757) 221-4386.
- For additional support or resources & questions, Contact the Dean of Students at 757-221-2510.

Important dates

date	event
2023-08-16	Last day to accept non-degree seeking applications
2023-08-25	New Student Orientation
2023-08-26	New Student Orientation
2023-08-27	New Student Orientation
2023-08-28	New Student Orientation
2023-08-29	New Student Orientation
2023-08-29	Add/drop period begins at 1:00 p.m.

date	event
2023-08-30	First day of classes Non-degree seeking registration begins
2023-09-04	Labor Day (offices closed, no classes)
2023-09-11	Last day to add/drop
2023-09-12	Withdrawal period begins UG Pass/Fail
2023-10-06	Deadline for 2024 graduates to have social class changed
2023-10-09	Midterm grading period
2023-10-10	Midterm grading period
2023-10-11	Midterm grading period
2023-10-12	Midterm grading period
2023-10-12	Fall Break
2023-10-13	Midterm grading period
2023-10-13	Fall Break
2023-10-14	Midterm grading period
2023-10-14	Fall Break
2023-10-15	Midterm grading period
2023-10-15	Fall Break
2023-10-16	Midterm grading period
2023-10-16	Classes resume from Fall Break
2023-10-17	Midterm grading period
2023-10-18	Midterm grading period
2023-10-19	Midterm grading period
2023-10-20	Midterm grading period
2023-10-21	Midterm grading period
2023-10-22	Midterm grading period
2023-10-23	Midterm grading period
2023-10-23	Advising period

date	event
2023-10-24	Midterm grading period
2023-10-24	Advising period
2023-10-25	Midterm grading period
2023-10-25	Advising period
2023-10-26	Midterm grading period
2023-10-26	Advising period
2023-10-27	Midterm grading period
2023-10-27	Advising period
2023-10-28	Midterm grading period
2023-10-28	Advising period
2023-10-29	Midterm grading period
2023-10-29	Advising period
2023-10-30	Advising period
2023-10-30	Last day to withdraw from a full-term course
2023-10-31	Advising period
2023-11-01	Advising period
2023-11-02	Advising period
2023-11-03	Advising period
2023-11-03	Deadline to declare major/minor
2023-11-07	Election Day (offices closed, no classes)
2023-11-13	Spring 2024 priority registration for continuing students
2023-11-14	Spring 2024 priority registration for continuing students
2023-11-15	Spring 2024 priority registration for continuing students
2023-11-16	Spring 2024 priority registration for continuing students
2023-11-20	Remote instruction days (courses in session remotely, offices closed)
2023-11-21	Remote instruction days (courses in session remotely, offices closed)

date	event
2023-11-22	Thanksgiving Break (offices closed, no classes)
2023-11-23	Thanksgiving Break (offices closed, no classes)
2023-11-24	Thanksgiving Break (offices closed, no classes)
2023-11-25	Thanksgiving Break (offices closed, no classes)
2023-11-26	Thanksgiving Break (offices closed, no classes)
2023-11-27	Classes resume from Thanksgiving Break
2023-12-08	Last day of classes
2023-12-09	Reading periods
2023-12-10	Reading periods
2023-12-11	Final exam periods
2023-12-12	Final exam periods
2023-12-13	Final exam periods
2023-12-14	Final exam periods
2023-12-15	Final exam periods
2023-12-16	Reading periods
2023-12-17	Reading periods
2023-12-18	Final exam periods
2023-12-19	Final exam periods

- Web

W&M honor code

Students are expected to conduct themselves according to the Honor Code.