COMP8730
Natural Language Processing and Understanding
WINTER 2022
COURSE SYLLABUS

SCHOOL OF COMPUTER SCIENCE

LAND ACKNOWLEDGEMENT

The University of Windsor sits on the traditional territory of the Three Fires Confederacy of First Nations, which includes the Ojibwa, the Odawa, and the Potawatomi. We respect the longstanding relationships with First Nations people in this place in the 100-mile Windsor-Essex peninsula and the straits – les détroits – of Detroit

INSTRUCTOR:

Hossein Fani

E-mail: hfani@uwindsor.ca
Office Location: *Lambton Tower 5111, Online (Blackboard)*
Office Hours: **Mondays 02:30PM-03:30PM**

*Note: Only email originating from a valid University of Windsor student account will be accepted from students wishing to contact the instructor or use the Blackboard email tool within the course site. Please include your full name, student ID and related course section in your correspondence. Do not spam with multiple or lengthy emails. Should you not receive timely feedback to your inquiries reach out during office hours directly, or in the event of no response contact the CS office at **csinfo@uwindsor.ca** for support to access the instructor.*

*The course outline that is available after the end of second week of semester will be deemed correct and official. *

Never used Microsoft Teams before?
Download the free MS-Teams client for your device and login using your UWindsor account (uwinid). There are two ways to reach me, one using the direct chat to Hossein Fani and another to our class group if you like to connect with your peers. It is a simple messenger type application allowing you to do chat, voice and video conferences with your prof and fellow students.
Getting Started - Students | Information Technology Services (uwindsor.ca)

PRE-REQUISITES: COMP-8700: Introduction to Artificial Intelligence or Permission of the Instructor

LECTURES Mondays 11:30AM-02:20PM, Dillon Hall 352 (Hyflex)

COURSE DESCRIPTION*:

This course covers the basic linguistic, logical and AI approaches to the development of natural language understanding systems. Topics covered include: syntactic/parsing strategies, formal semantics, pragmatics and the resolution of various types of ambiguities. Inference strategies involved in the resolution of ambiguities at the pragmatic level include a detailed discussion of the representation of and reasoning with common sense knowledge. The course also includes the implementation of natural language interfaces and the application of linguistic approaches to the development of intelligent text retrieval systems.
**This description is from the official senate-approved calendar*

LEARNING OUTCOMES:

At the end of the course, the successful student will know and be able to:

A. The acquisition, application and integration of knowledge
Explain natural language processing methods based on textual content for:
- text classification,
- text summarization,
- text clustering,
- sentiment analysis,
- topic modelling,
- query-based information retrieval
- information extraction

and apply them to concrete, practical problems.

B. Research skills, including the ability to define problems and access, retrieve and evaluate information (information literacy):
Define natural language processing methods based on textual content for solving:
- text classification,
- text summarization,
- text clustering,
- sentiment analysis,
- topic modelling,
- query-based information retrieval
- information extraction

problems.

C. Critical thinking and problem-solving skills
- Solve practical problems by using and combining several natural language processing concepts or strategies.

D. Literacy and numeracy skills
- Write coherently, concisely, and clearly in a range of formats and for a variety of audiences.

E. Responsible behaviour to self, others and society
- Explain the societal, legal and ethical implications of natural language processing methods and solutions.

F. Interpersonal and communications skills
- Present research work to computer science audiences and answer questions.

G. Teamwork, and personal and group leadership skills: N/A

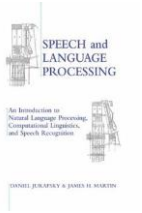
H. Creativity and aesthetic appreciation
- Design new natural language processing approaches.

I. The ability and desire for continuous learning
- Identify how good theoretical and practical knowledge of problems leads to efficient natural language processing and understanding solutions.

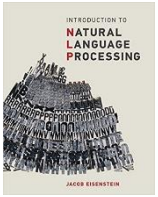
Note: Students are strongly encouraged in participating in the course development and update process. Please feel free to make recommendations for changes of the Learning Outcomes, Course Description, and Course Topics to the instructor or the program chair.

REQUIRED TEXTBOOK:

Speech and Language Processing, 3rd Ed. Draft
An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition
Dan Jurafsky and James H. Martin
Free Access → https://web.stanford.edu/~jurafsky/slp3/



Introduction to Natural Language Processing
Jacob Eisenstein
ISBN: 9780262042840
October 2019
Free Access → http://cseweb.ucsd.edu/~nnakashole/teaching/eisenstein-nov18.pdf



- *Campus Bookstore: https://www.uwindsor.ca/bookstore/*
- *Leddy Library: https://leddy.uwindsor.ca/*

COURSE EVALUATION:

Research Project - Problem Definition - Literature Review - Presentation I: Proposal - Proposed Solution - Experiment - Presentation II: Demo - Peer Review - Conference Submission (Pending Instr. 's Approval)	70% - 05% - 10% - 05% - 10% - 20% - 05% - 05% - 10%	
Assignments	10%	
Midterm Exam	10%	Feb 28, 2022
Final Exam	10%	Apr. 20 – 29

COURSE SCHEDULE:

Topics*

(The instructor reserves the right to change the outline to accommodate student pace and understanding of the subject matter.)

Lec01: Meet & Greet, Introduction	Jan. 17
Lec02: Text Normalization	Jan. 24
Lec03: n-Gram Language Models	Jan. 31
Lec04: Sentiment Classification	Feb. 07
Lec05: Logistic Regression	Feb. 14
Lec06: **Reading Week (No Classes)**	Feb. 19 – Feb. 27
Lec07A: **Midterm** Lec07B: Project Presentation I: Proposal	Feb. 28 Feb. 28
Lec08: Vector Semantics and Embedding	Mar. 04
Lec09: Neural Language Models	Mar. 14
Lec10: Sequence Labeling	Mar. 21
Lec11: Topic Modeling, User Modeling	Mar. 28
Lec12: Web Search	Apr. 04
Lec13: Project Presentation II: Demo	Apr. 11
Lec14: Final Exam *Date and Time to be Announced.*	Apr. 20 – 29

***Note:** Students are advised that the schedule and topics described above are tentative and that the material and/or depth and order of presentation are subject to change at the discretion of the instructor and student pace.*
This course assumes the student will allocate a significant amount of independent study and time spent on reading and researching materials as needed. You are strongly encouraged to ensure sufficient time needed to succeed in this course.

PROJECT SCHEDULE:

This course is research-oriented and project-driven:

- Your knowledge will be mostly evaluated based on an ongoing research project defined in the NLP-related area.
- The research project can be selected from the available research problems or should be approved by the instructor.
- Each milestone is accompanied with a manual that explains the task and submission procedure as well the deadline.
- Project assignments are expected to be submitted on the assigned due date and time.
- Late submission receives 0 unless a verifiable reason with appropriate documentation is provided.
- The students should follow the submission procedure for each project assignment. Failure to follow the procedure (e.g., incorrect, unreadable, and missing file attachments as instructed) heavily penalizes the submission.
- Each assignment must be done individually or in a team of maximum 2 students.
- The research project's milestones are interrelated and are met in order.
- Conference submission needs the instructor approval pending the quality of the research manuscript meets the high standards for scientific scholarly papers!
- ***Bonus**: In the case of conference acceptance of your research paper, the instructor will cover the conference registration fee. Also, the instructor will try to fund the conference travel for 1 student.*

1: Proposal (Project Definition)	Authors (Project Team Members, Max: 2) Abstract Motivation Application Formal Definition (Math)
2: Literature Review	History [-∞, 2010] State of the Art (SOTA) [2010, Present] Critics and Gaps What Gaps the Project Fills?
3: Presentation I: Proposal	10 Minutes Presentation 05 Minutes QA
4: Proposed Solution	Formal Definition Pipeline Steps

	Implementation (Online Repo, e.g., Github)
5: Experiments	Evaluation Methodology Metrics Datasets Baselines Results
6: Presentation II: Demo	10 Minutes Presentation 05 Minutes QA
7: Peer Review	
8: Conference Submission	Pending Instructor Approval
9: Conference Acceptance	Bonus → Conference Registration and Travel

RESOURCES:

The course website is on ***https://blackboard.uwindsor.ca/***
Please check it frequently for announcements and other useful info.

GRADING:

A numeric grade on a scale of 0 to 100 will be assigned (rounded integer).

Passing grade:
A minimum grade of 70% is required to pass this course. Your individual program may have higher requirements to maintain good standing; please consult your program requirements and plan accordingly. If you are registered in a course and do not attend or participate or write any evaluations will be assigned a grade of NR (No report). You must withdraw from the course if you do not wish to attend it; not showing up does not constitute withdrawal and will impact your academic record.

Voluntary withdrawal (dropping the course):
You may drop a course within the first 2 weeks add/drop period (1 week in case of 6-week courses) without it showing up on your academic record. Please check with the Registrar's office calendar on the important dates for withdrawing voluntarily from a course after the add/drop period should you feel you need to withdraw. It is strongly recommended that you seek academic advice from your instructor or an academic advisor prior to withdrawing from courses.

Absences due to medical or other extenuating circumstances:
Medical leaves, illness, death (in the family), and other difficult circumstances as determined in bylaw 54 are at times unavoidable and would interrupt your academic career. You must report any issues to the instructor as soon as possible prior to considering any academic accommodations. The instructor reserves the right to determine if an accommodation is merited and the nature of the accommodation related to the course evaluation. All requests for alternate considerations on medical grounds or other difficult matters must be made in writing (email) to the instructor along with supporting documents prior to the end of the course. No alternate accommodations will be considered after the end of the course.

Makeup and missed assessment policy:
If you miss a test, assignment or other assessment in the course you will receive a zero mark for the missed work. If you wish to have alternate considerations due to a valid reason (as per senate bylaw 54) you must inform the instructor in writing (email) as soon as possible, preferably before the assessment, and not later than seven calendar days. Considerations for any make-up or late submissions will be done on a case-by-case basis on compassionate grounds while maintaining fairness as much as possible. No alternate considerations will be given to any missed assessment if the instructor is not informed within seven calendar days after its due date. The instructor will refuse any unsubstantiated and late requests.

- ***Midterm Exam:*** *should a student miss a midterm exam, with appropriate documentation and verifiable reason, the weight of the missed midterm exam will be moved to the final exam.*
- ***Final Exam: s****tudents who miss a final exam for a verifiable reason and with appropriate documentation will be given a make-up exam prior to the submission of final course grades that carries the same weight and measure the same knowledge.*
- ***Make-up of the Make-up*****:** *there will be no make-up of the make-up exam, and the final grade will be assigned based on overall work.*
- ***Late Submission:*** *late submission receives zero unless a verifiable reason with appropriate documentation is provided.*

Grade appeal:
Informal reviews and appeals of the marks for assignments, midterm, exams and/or projects will be considered only if requested within 10 days after the release of the corresponding grades. After the 10-day period students will have to submit a formal appeal if they wish within 6 weeks. See Senate Bylaws 54 (Undergraduate Students) and Senate Bylaws 55 (Graduate Students) for more details on appealing about grades.

Other Notes:
1.A. Undergraduate Students: (Please review Bylaw 54) The last seven calendar days prior to, and including, the last day of classes are free from any procedures for which a mark will be assigned. (Extensions on compassionate grounds are excluded). (In the case six weeks courses, the last three calendar days before the start of the examination period are free from any assessment procedures).
1.B. Unannounced quizzes/graded activities will not exceed 5% of the final grade.
1.C. Participation marks in online courses will not exceed 20% of the final grade.

2. The final exam schedule is announced by the Registrar's office, normally after the add/drop period, and students are expected to be available for the entire exam period and not make any prior travel plans, vacations, or other commitments until after the exam dates are announced. No alternate exams accommodations will be made on those grounds.

3. No forms of assessment shall be scheduled or made-due on days identified as break days such as reading weeks, holidays, or days that the University is officially closed.

SET:

Student Evaluation of Teaching (SET) will be administered in the last 2 weeks of classes via UWinsite (or last week of classes in the case of 6-week classes) as per Senate policy.

SUPPORT CONTACTS:

The School of Computer Science has a team of support staff and access to student academic advisors to assist you through any inquiries you may have about our courses and programs. Please use one of the following emails:
For CompSci undergraduate programs and advising, including IT certificate: csinfo@uwindsor.ca
For CS Tutors (free tutoring support for all CS undergrad courses): http://tutor.cs.uwindsor.ca/
For Computer Science Society: https://css.uwindsor.ca/
For CompSci graduate programs (MSc, MSc-AI stream, and PhD): csgradinfo@uwindsor.ca
For CompSci professional graduate programs (MAC/MAC-AI stream): macprogram@uwindsor.ca
For the office of the Director of the School of Computer Science: csdir@uwindsor.ca
For CompSci technical support: https://help.cs.uwindsor.ca/
For International Student Centre: https://www.uwindsor.ca/international-student-centre/
For Student Accessibility Services: https://www.uwindsor.ca/studentaccessibility/
For other general inquiries: https://ask.uwindsor.ca/
For student counselling services (ext. 4616): https://www.uwindsor.ca/studentcounselling/
For student health services (ext. 7002): https://www.uwindsor.ca/studenthealthservices/
For student Peer Support Centre (ext. 4551): https://www.uwindsor.ca/studentexperience/wellness/
For USci Faculty of Science student support network: https://www.uwindsor.ca/science/usci/

Need help?
My Student Support Program (MySSP) is an immediate and fully confidential 24/7 mental health support that can be accessed for free through chat, online, and telephone. This service is available to all University of Windsor students and offered in over 30 languages.
Call: ***1-844-451-9700****, or visit* ***https://myissp.com/***

STUDENT ACCOMMODATIONS:

Students with disability:
Students who require academic accommodations in this course due to a documented disability must contact an Advisor in Student Accessibility Services (SAS) to complete SAS Registration and receive the necessary Letters of Accommodation. After registering with SAS, you must present your Letter of Accommodation and discuss your needs with the course instructor as early in the term as possible. Please note that deadlines for the submission of documentation and completed forms to SAS are available on their website:

- http://www.uwindsor.ca/studentaccessibility/

Exam conflicts:
If you have a conflict with two exams at the same time, you will need to talk to both instructors and ask which one is willing to move your exam to a different day or time.

If you have a conflict with examinations due to the following reasons, view the Office of Registrar Alternative Final Exam Policy:

- Conflict with religious conviction during the regularly scheduled time slot.
- Three or more final examinations in a 24-hour period.

Religious Observances:
Requests for accommodation of specific religious or spiritual observance must be presented to the instructor no later than 2 weeks prior to the conflict in question (in the case of final examinations within two weeks of the release of the examination schedule). In extenuating circumstances, this deadline may be extended. If the dates are not known well in advance because they are linked to other conditions, requests should be submitted as soon as possible in advance of the

required observance. Timely requests will prevent difficulties in arranging constructive accommodations. religious_accommodation_for_students.01mar2013.web_ver.pdf (uwindsor.ca)

PRIVACY AND COPYRIGHTS:

Content confidentiality:
Lectures, examinations, quizzes, assignments, and projects given in this course are protected by copyright. Reproduction or dissemination of examinations or the contents or format of examinations/quizzes in any manner whatsoever (e.g., sharing content with other students or websites), without the express permission of the instructor, is strictly prohibited. Students who violate this rule or engage in any other form of academic dishonesty will be subject to disciplinary action under Senate Bylaw 31: Student Affairs and Integrity.

Recording of lectures:
Lectures and discussions can be recorded by requesting explicit permission from the instructor. Students planning to do so shall send a request (via email is sufficient) before the lecture is delivered. Students, however, are not allowed to post or share any recorded material to any other individual or party outside of this course.
See Senate Policy on recording lectures.

SAFETY, ACADEMIC INTEGRITY, AND NON-ACADEMIC MISCONDUCT:

Equity, Diversity, and Inclusiveness (EDI)
This course, along with all its components such as lab sections are, without question, safe places for students of all races, genders, sexes, ages, sexual orientations, religions, disabilities, and socioeconomic statuses. Disrespectful attitude, sarcastic comments, offensive language, or language that could be translated as offensive and/or marginalize anyone are absolutely unacceptable. Immediate actions will be taken by the instructor to protect the safety and comfort of the students. An ethnically rich and diverse multi-cultural world should be celebrated in the classroom. The instructor, too, must treat every student equally and with the respect and compassion that all students deserve. Furthermore, UWindsor is committed to combatting sexual misconduct. All members are required to report any instances of sexual misconduct, including harassment and sexual violence, to the Sexual Misconduct Response & Prevention Office so that the victim may be provided appropriate resources and support options.

- https://www.uwindsor.ca/sexual-assault/
- For police/ambulance emergency call 911 (in Canada)
- For campus police call 519-253-3000 ext. 4444 for emergency, and 1234 for non-emergency issues.

Academic Integrity
Please refer to: https://www.uwindsor.ca/academic-integrity/

As defined in the University of Windsor's **Student Code of Conduct**, plagiarism is the act of copying, reproducing or paraphrasing significant portions of one's own work, or someone else's published or unpublished material (from any source, including the internet), without proper acknowledgement, representing these as new or as one's own.

Tips and resources to help you prevent plagiarism:
https://www.uwindsor.ca/academic-integrity/sites/uwindsor.ca.academic-integrity/files/tips_for_preventing_plagiarism.pdf

The instructor will put a great deal of effort into helping students to understand and learn the material in the course. However, the instructor will not tolerate any form of cheating. The instructor will report any suspicion of academic integrity to the Director of the School of Computer Science. If sufficient evidence is available, the Director will begin a formal process according to the University Senate Bylaws which will lead to more review, a strict punishment if convicted, and a note on your permanent student record.

The following behaviours will be regarded as cheating:

- *Copying assignments or quizzes or presenting someone else's work as your own.*
- *Allowing another student to copy an assignment/project from you and present it as their own work; protect your own work and never share it with anyone!*
- *Copying from another student or any other unauthorized source during a test or exam.*
- *Falsifying your identity during the exam or having someone else assist or complete your assessment.*
- *Referring to notes, textbooks, and any unauthorized sources during a test or exam (unless otherwise stated).*
- *Speaking or communicating without permission during a test or exam.*
- *Not sitting at the pre-assigned seat during a test or exam.*
- *Communicating with another student in any way during a test or exam.*
- *Having unauthorized access to the exam/test paper prior to the exam/test.*
- *Explicitly asking a proctor for the answer to a question during an exam/test.*
- *Modifying answers after they have been marked.*
- *Any other behaviour which attempts unfairly to give you some advantage over other students during the grade-assessment process.*
- *Refusing to obey the instructions of the officer in charge of an examination.*

The list given above is not exhaustive. More examples are given in Appendix A, Senate Bylaws 31 – Complete guidelines and procedures on the sanctions imposed by the university are also listed in Table A.1 of the Senate Bylaws 31

In this course any assessment that is deemed plagiarized or in violation of the academic integrity policy will NOT BE GRADED and receive a grade of ZERO unless a different ruling is provided by the adjudication committee formally reviewing the case.

Examples of sanctioning include: *(from Table A.1 in Appendix A of Bylaw 31)*
For first offence: mark reduction up to zero, censure 6-12 months; and for subsequent offence: suspension 4-24 months, censure up until graduation.

Plagiarism detection software:
Plagiarism-detection software *SafeAssign* will be used for all student assignments in this course. You will be advised how to submit your assignments. Note that students' assignments that are submitted to the plagiarism-detection software become part of the institutional database. This assists in protecting your intellectual property. However, you also have the right to request that your assignment(s) not be run through the student assignments database. If you choose to do so, that request must be communicated to the course instructor in writing at the beginning of the course. The instructor reserves the right to choose another plagiarism detection software and students would be notified of this once it is put in use.