University of Alberta

CMPUT 497
Introduction to Natural Language Processing
Fall 2021

INSTRUCTORS

Name	Denilson Barbosa	Carrie Demmans Epp	Greg Kondrak
Preferred Name	Denilson	Carrie	Greg
Pronouns	He/him	She/her	He/him
Email	denilson@ualberta.ca	cdemmansepp@ualberta.ca	gkondrak@ualberta.ca

TEACHING ASSISTANTS
See eClass

OFFICE HOURS
See eClass

Lectures: Tuesdays and Thursdays 3:30PM - 4:50PM
via Zoom (check eClass -- must be logged to Zoom with your ualberta account)
2 Sept. 2021 – 7 Dec. 2021

All times in UTC -6 until 7 Nov. 2021. After 7 Nov. 2021 UTC -7 will be used.

COURSE DESCRIPTION
Natural language processing (NLP) is a subfield of artificial intelligence concerned with the interactions between computers and human languages. This course is an introduction to NLP, with the emphasis on writing programs to process and analyze text corpora. The course covers both foundational aspects and applications of NLP. The course aims at a balance between classical and statistical methods for NLP, including methods based on machine learning.

In this course, students will
- clean or otherwise pre-process natural language corpora that could include 1000s of documents
- develop natural language processing tools
- integrate existing tools into an analysis task
- apply computational methods to natural language artefacts to extract information, classify the language within the artefact, identify relationships among artefacts, or identify relationships among elements within an artefact

Course Prerequisites
CMPUT 201 or 275 and any 300-level CMPUT course. Beyond these specific prerequisites, it is expected students will have a strong background in Computer Science and can write programs independently. We also expect that students can install and use existing libraries or utilities.

While we are happy to recommend resources, students who do not have the required prerequisites at the time of taking this course should not expect members of the instructional team to provide supplementary professorial tutoring.

Course Objectives and Expected Learning Outcomes

Students will be able to

1. thoroughly examine and discuss the intricacies involved in working with natural language data through their selection, application, and justification of data cleaning and modelling procedures;
2. provide nuanced interpretations and contextualizations through discussion with others or when presenting their work-products (e.g., assignment reports)
3. compare and contrast classical algorithms and their statistical counterparts for several important applications.

COURSE SCHEDULE & READINGS

We expect you to come to the synchronous sessions prepared. The specifics of what is required in each week will be communicated to you on an on-going basis via eClass. The dates when topics will be covered may be adjusted to meet student needs (i.e., if students are struggling with a key concept we may take longer to cover it and the time dedicated to topics students seem to grasp quickly may be reduced).

Readings

These will be specified via eClass. They are generally available via the Internet and library. Those that cannot be freely obtained by University of Alberta students will be made available through eClass**.**

Be prepared to discuss course materials critically in view of the current course topic as well as your personal experiences and interests. Also read more broadly (as implied by the wiki that contains references to related resources) and by following up on publications cited in the articles under discussion. Please contribute to class discussions based on your current thinking, academic experiences, and professional experiences.

For individual queries feel free to use e-mail to contact the course instructor or other participants in the course, as appropriate. Other questions should be posted to the eClass Discussion Forum. Queries that we feel other students will benefit from will be de-identified and answered via the Discussion Forum.

Recommended or Optional Learning Resources

Additional resources and readings are also provided via eClass. They are meant to provide those who are interested with additional background information and help provide greater depth for course topics that relate to the project you select.

On-Line Coursework Disclaimer

On-line work is a component of this course. The systems you will use can be accessed freely and some are hosted by third-party organizations. Please be aware that these organizations may store information about your activities within their systems. You are not required to provide any personal information to these organizations; this includes your CCID and @ualberta email address.

In this course, we will be using GitHub. You may use an existing GitHub account, or you can create a new one. You can use your ualberta anonymous ID to create an account. Appending @ualberta.ca to your anonymous ID will create a valid email address that protects your identity while ensuring that emails will be delivered to your ualberta account.

See https://support.ctl.ualberta.ca/index.php?/Knowledgebase/Article/View/350/10/finding-your-anonymous-id-in-eclass for information about how to find your anonymous ID.

If you have any concerns about this, please contact the instructor of the course.

GRADE EVALUATION

Criterion-based grading will be used.

For assignments, students will be assessed in accordance with criteria specified in the assignment description and the rubric.

Your final grade will be based on our interpretation of the grading system as defined in the Calendar. We do not use a pre-defined function of your final mark to compute your final grade, but instead use our judgement of how the class final marks reflect mastery of the course material.

Assignments are normally done under the department's Full Collaboration model (see Course Policies). For assignments, your group can have at most two students. If you have collaborated with other students on your assignment, you must list their names in your submission. Only one of you has to submit the assignment. Also, you must list any other resources that you used to develop your solution.

Unless otherwise specified deliverables are due at 11:55PM Edmonton time

Deliverable	Weighting		Date
Quizzes (Best 5 of 6)	**15** %	(5 × 3%)	Bi-weekly
Assignments	**50** %	(5 × 10%)	see eClass
Final Exam	**35** %		TBA

Grades are unofficial until approved by the Department and/or Faculty offering the course.

Evaluative Material

Please see eClass for samples of past evaluative course materials and rubrics (grading criteria).

Format of Exams

Quizzes and exams may include

- *Application of NLP techniques*
- *Comprehension & synthesis questions*
- *Analysis of text*
- *Design of an NLP technique*

Missed Course Work

For an excused absence where the cause is religious belief, a student must contact the instructor(s) within two weeks of the start of Fall or Winter classes to request accommodation for the term (including the final exam, where relevant). Instructors may request adequate documentation to substantiate the student request.

A student who cannot write a term examination or complete a term assignment due to incapacitating illness, severe domestic affliction or other compelling reasons can apply for an extension of time to complete an assignment. Please contact the instructor via email if you require an extension or other accommodation. In all cases, instructors may request adequate documentation to substantiate the reason for the absence at their discretion.

They can also apply for a deadline extension to complete an assignment or laboratory should the time required exceed their late credits (see below).

Late credits and penalties

You are expected to complete assignments on time. There will be a penalty for lateness of 10% deducted per day or part thereof. Work that is not handed within 2 days of the advertised due date will not be accepted. If you require more time to complete term work you should contact your instructor immediately, and no later than the due date. With the exception of quizzes and exams all deliverables are eligible for late submission.

Late submission grace credits

Each student receives a 4-day "no questions asked" credit that can be applied, in 1-day increments, to eligible deliverables (assignments and project milestones). This is automatically deducted based on the time stamp of the submission. To use these credits, you must email Carrie (demmanse@ualberta.ca) before the assignment deadline.

Deferral of term work is a privilege and not a right; there is no guarantee that a deferral will be granted. Misrepresentation of Facts to gain a deferral is a serious breach of the *Code of Student Behaviour*.

DEFERRED FINAL EXAMINATION

A student who cannot write the final examination due to incapacitating illness, severe domestic affliction or other compelling reasons can <u>apply</u> for a deferred final examination. Students who failed at the start of term to request exam accommodations for religious beliefs are expected to follow the normal deferred final examination process. Such an application must be made to the student's Faculty office within two working days of the missed examination and must be supported by appropriate documentation or a Statutory Declaration (http://calendar.ualberta.ca/content.php?catoid=20&navoid=4939#Attendance). Deferred examinations are a privilege and not a right; there is no guarantee that a deferred examination will be granted. Misrepresentation of Facts to gain a deferred examination is a serious breach of the *Code of Student Behaviour*.

The deferred final examination will be held **10 January 2022 at 9:00am (UTC -7)**. All deferred exams will be oral exams.

REMOTE DELIVERY CONSIDERATIONS

Until further notice, the course will be offered "in Person". Instructors will attempt to accommodate students who cannot or prefer not to attend in person by broadcasting and recording lectures and by conducting some activities online, to the extent possible.

Technology for Remote Learning

To successfully participate in remote learning in this course, it is recommended that students have access to a computer with an internet connection that can support the tools and technologies the University uses to deliver content, engage with instructors, TAs, and fellow students, and facilitate assessment and examinations. Please refer to [Technology for Remote Learning - For Students](#) for details. If you encounter difficulty meeting the technology recommendations, please email the Dean of Students Office (dosdean@ualberta.ca) directly to explore options and support.

Please contact the instructor before the add/drop deadline (**15 September 2021**) if you do not have access to the minimum technology recommended. The instructor will make arrangements for accommodating students who contact the instructor before this date. Failure to do so may result in a zero in any assessment that depends on the minimum technology.

Recordings of Synchronous Activities
- Please note that class times for this course will be recorded. Recordings of this course will be used for the purposes of documentation and to ensure that everyone obtains access to all course materials. Recordings will be disclosed to other students enrolled in this section of the class and members of the instructional team.

- Students have the right to not participate in the recording and are advised to turn off their cameras and audio prior to recording; they can still participate through text-based chat. It is recommended that students remove all identifiable and personal belongings from the space in which they will be participating.

- Recordings will be made available until the beginning of the next term and accessible via Google Drive. Please direct any questions about this collection to the instructor of this course.

Home-Based Lab Activities
As part of the learning experience in this course, you will be required to undertake certain activities in or around your place of residence. To ensure that you undertake the activity safely and are fully informed of the risks, please review the University of Alberta's [Remote Learning Lab-Based Assignment Information Advisory](). If you have questions or concerns, contact your instructor.

Student Resources for Remote Learning
Online learning may be new to you. Check out tips for success and find out more about online learning on the [Campus Life]() page, and specifically on the [Student Resources for Remote Learning]() page.

COURSE POLICIES
This course is subject to the [Department of Computing Science policies](). In particular, pay attention to these [Computing Science Course Policies]().

Exam Conduct
- Your student photo I.D. is required at exams to verify your identity.
- Students will not be allowed to begin an examination after it has been in progress for 30 minutes. Students must remain in the exam room until at least 30 minutes has elapsed.
- All cell phones must be turned off and stored in your bags.
- *See http://calendar.ualberta.ca/content.php?catoid=20&navoid=4939 - Examinations_(Exams) for additional information about exam procedures.*

Cell Phones
Cell phones are to be turned off or otherwise silent during lectures, labs, and seminars.

Students eligible for accessibility-related accommodations (students registered with Student Accessibility Services - SAS)
Eligible students have both rights and responsibilities with regard to accessibility-related accommodations. Consequently, scheduling exam accommodations in accordance with SAS deadlines and procedures is essential. Please note adherence to procedures and deadlines is required for U of A to provide accommodations. Contact SAS (www.ssds.ualberta.ca) for further information.

Student Success Centre
Students who require additional help in developing strategies for better time management, study skills or examination skills should contact the Student Success Centre (2-300 Students' Union Building).

Recording and/or Distribution of Course Materials
Audio or video recording, digital or otherwise, of lectures, labs, seminars, or any other teaching environment by students is allowed only with the prior written consent of the instructor or as a part of an approved accommodation plan. Student or instructor content, digital or otherwise, created and/or used within the context of the course is to be used solely for personal study, and is not to be used or distributed for any other purpose without prior written consent from the content author(s).

6

Policy about course outlines can be found in the University Calendar
([http://calendar.ualberta.ca/content.php?catoid=6&navoid=806&hl="syllabus"&returnto=search - Evaluation_Procedures_and_Grading_System](http://calendar.ualberta.ca/content.php?catoid=6&navoid=806&hl="syllabus"&returnto=search)).

DISCLAIMER
Any typographical errors in this Course Outline are subject to change and will be announced in class.

ACADEMIC INTEGRITY
"The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behaviour and avoid any behaviour which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University."

All forms of dishonesty are unacceptable at the University. Any offence will be reported to the Associate Dean of Science who will determine the disciplinary action to be taken. Cheating, plagiarism, and misrepresentation of facts are serious offences. Anyone who engages in these practices will receive at minimum a grade of zero for the exam or paper in question and no opportunity will be given to replace the grade or redistribute the weights. As well, in the Faculty of Science the sanction for **cheating** on any examination will include **a disciplinary failing grade** (NO EXCEPTIONS) and senior students should expect a period of suspension or expulsion from the University of Alberta.

See these resources for information and examples on what may or may not be acceptable:

- https://cloudfront.ualberta.ca/-/media/science/research-and-teaching/documents/2017/dont-do-it-august-2017.pdf
- https://www.ualberta.ca/centre-for-teaching-and-learning/grants/tlef/tlef-deliverables/academic-integrity
- https://cloudfront.ualberta.ca/-/media/science/research-and-teaching/documents/sciencelabcoursehallofshame.pdf

COPYRIGHT
Drs. Demmans Epp, Barbosa, and Kondrak, Department of Computing Science, Faculty of Science, University of Alberta (2021).