

CS 4417/9647/9917b: Unstructured Data Course

The University of Western Ontario

Course Outline – Spring 2021

1. General Course Information

Course Information

CS 4417b: Unstructured Data

Tuesday: 4:30-5:30; Thursday 2:30-4:30

The course is primarily asynchronous but synchronous activities (e.g. tests and tutorials) may occur during scheduled class times.

Students should refer to the course OWL site for all scheduling information.

Prerequisite Requirements

This course requires Computer Science 3319 A/B.

Unless you have either the requisites for this course or written special permission from your Dean to enroll in it, you may be removed from this course and it will be deleted from your record. This decision may not be appealed. You will receive no adjustment to your fees in the event that you are dropped from a course for failing to have the necessary prerequisites.

2. Instructor Information

This course is being co-taught by two instructors.

- Dan Lizotte – dlizotte at uwo.ca – MC 363
- Hanan Lutfiyya – hlutfiyy at uwo.ca – MC 355

Students must use their Western (@uwo.ca) email addresses when contacting their instructors.

Office hours/Q&A will be held during the Thursday 2:30pm to 4:30pm timeslot on Zoom.

3. Course Description

This course will be delivered asynchronously via OWL and synchronously via Zoom. Asynchronous lecture content will be posted on OWL on Mondays, and the 2:30pm to 4:30pm timeslot on Tuesdays will be available for students to ask questions about the course content and assignments. Additional Zoom meetings (e.g. for midterm or exam review) may be scheduled during class time.

The objective of this course is to introduce students to techniques for the management and analysis of unstructured data, with a focus on text data e.g., transaction logs, news text, article

abstracts, and microblogs. The course will also provide an overview of unstructured image, audio, and video data. Students will receive hands-on experience with modern distributed data management and analysis infrastructure.

On successful completion of this course, students shall be able to:

- Contrast unstructured data source and management solutions with structured ones
- Recognize when unstructured data sources can support a data analytics solution
- Estimate the computation and storage needs for analyzing and storing a given large unstructured data source
- Use a modern unstructured data management tools, e.g., MongoDB or Hadoop to store and process a large unstructured data source
- Select, apply and evaluate appropriate analysis methods for unstructured text data
- Explain how the analysis methods work and explain their results to educated non-experts

4. Course Materials

Suggested Textbooks

The following textbooks are recommended.

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.

Online: <https://nlp.stanford.edu/IR-book/>

Stephan Buttcher, Charles L.A. Clarke, Gordon V. Cormack, *Information Retrieval: Implementing and Evaluating Search Engines*, MIT Press, 2010.

Online: <https://mitmecsept.files.wordpress.com/2018/05/stefan-bc3bcttcher-charles-l-a-clarke-gordon-v-cormack-information-retrieval-implementing-and-evaluating-search-engines-2010-mit.pdf>

Additional references specific to a topic may be provided.

Lecture Notes

Lecture notes will be made available in OWL.

5. Methods of Evaluation

The overall course grade will be calculated as listed below:

CS4417 Students

Assignments (3)	50%
Midterm Exam	25%
Final Exam	25%

CS9647/CS9117 Students

Assignments (3)	40%
Midterm Exam	25%

Final Exam	25%
Technical Topic Report	10%

The following rules are designed to ensure that students meet certain minimum standards:

- To be eligible to pass the course, a student must receive at least 40% in the exams and at least 40% in the assignments.
- To be eligible to receive an overall grade of 60% or higher in the course, a student must receive at least 50% in the exams and at least 50% in the assignments.

Exams

The dates for the exams are below:

- **Midterm exam:** February 25th in class
- **Final Exam:** Scheduled by the Registrar.

There will be no makeup midterm exam, except for students requesting a special midterm exam for religious reasons. These students must have notified the course instructors and filed documentation with their Dean's office at least two weeks prior to the midterm exam. If you miss the midterm exam for any other reason, follow the procedure for Academic Accommodation for Medical Illness. If accommodation is approved by your Dean's office, the final exam will be reweighted to include the weight of the midterm exam.

Assignments.

- Assignments will be made available on OWL. Students are responsible for checking the course web page on a regular basis. Submission instructions will be provided for each assignment
- There will be three assignments (worth 50% of the final mark) with the following breakdown:

Assignment	Due Date	Expected Load	Weight
1	February 12	Medium	14%
2	March 19	High	18%
3	April 9	High	18%

- The student evaluation procedure cannot be changed from that given in the course outline without prior consultation with the Undergraduate Chair and approval of the Dean of Science.
- All assignments are due by 11:55pm (electronically) of the due date. Late assignments may be handed in, but with a penalty of 5 marks a day. Saturday and Sunday together count as one day. Assignments more than 5 days late will not be accepted.
- To guard against the possibility of lost assignments and errors in recording marks, students are advised to retain all assignments that are handed back to them, as well as disk copies of all materials they hand in, at least until final marks for the course are posted.

Technical Topic Report

Graduate (CS 9637 and CS 9117) students will prepare a brief report on a new technical topic related to the management and/or analysis of unstructured data that interests them and that they feel would make a good addition to the course. Details on the structure of the brief report will be provided in class.

6. Accommodation and Accessibility

Accommodation Policies

Students with disabilities work with Accessible Education (formerly SSD) which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The Academic Accommodation for Students with Disabilities policy can be found at:

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Accommodation_disabilities.pdf

Academic Consideration for Student Absence

Students will have up to two (2) opportunities during the regular academic year to use an on-line portal to self-report an absence during the semester, provided the following conditions are met: the absence is no more than 48 hours in duration, and the assessment for which consideration is being sought is worth 30% or less of the student's final grade.

Any student who self-reports an absence for an assignment will be given a 48 hour extension.

Any student who self-reports an absence for the midterm will have the weight of the final exam adjusted to accommodate the midterm.

Please note that when you use an SRA we automatically receive an email reporting the SRA. In this communication, the student is asked to contact the instructor to arrange the accommodation. *We will not respond to these emails since the accommodation is described in the course outline.*

Students are not able to use the self-reporting option in the following circumstances:

- for exams scheduled by the Office of the Registrar (e.g., December and April exams)
- absence of a duration greater than 48 hours,
- assessments worth more than 30% of the student's final grade,
- if a student has already used the self-reporting portal twice during the academic year

If the conditions for a Self-Reported Absence are *not* met, students will need to provide a Student Medical Certificate if the absence is medical, or provide appropriate documentation if there are compassionate grounds for the absence in question. Students are encouraged to contact their Faculty academic counselling office to obtain more information about the relevant documentation.

Students should also note that individual instructors are not permitted to receive documentation directly from a student, whether in support of an application for consideration on medical grounds, or for other reasons. **All documentation required for absences that are not covered by the Self-Reported Absence Policy must be submitted to the Academic Counselling office of a student's Home Faculty.**

For policy on Academic Consideration for Student Absences - Undergraduate Students in First Entry Programs, see:

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Consideration_for_absences.pdf

and for the Student Medical Certificate (SMC), see:

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/medicalform.pdf

Religious Accommodation

Students should consult the University's list of recognized religious holidays, and should give reasonable notice in writing, prior to the holiday, to the Instructor and an Academic Counsellor if their course requirements will be affected by a religious observance. Additional information is given in the Western Multicultural Calendar:

<https://multiculturalcalendar.com/ecal/index.php?s=c-univwo>

You may also be eligible to write the Special Exam if you are in a “Multiple Exam Situation” (see http://www.registrar.uwo.ca/examinations/exam_schedule.html).

[optional] If a student fails to write a scheduled Special Examination, the date of the next Special Examination (if granted) normally will be the scheduled date for the final exam the next time this course is offered. The maximum course load for that term will be reduced by the credit of the course(s) for which the final examination has been deferred. See Academic Calendar for details (under [Special Examinations](#)).

7. Academic Policies

The website for Registration Services is <http://www.registrar.uwo.ca>.

In accordance with policy, <http://www.uwo.ca/its/identity/activatenonstudent.html>, the centrally administered e-mail account provided to students will be considered the individual's official university e-mail address. It is the responsibility of the account holder to ensure that e-mail received from the University at his/her official university address is attended to in a timely manner.

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at this website: http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf.

All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system.

Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).

Computer-marked, multiple-choice tests and exams may be subject to submission for similarity review by software that will check for unusual coincidences in answer patterns that may indicate cheating.

Submitted code may be subject to submission for similarity review by the MOSS software.

8. Support Services

Please visit the Science & Basic Medical Sciences Academic Counselling webpage for information on add/drop courses, academic considerations for absences, appeals, exam conflicts, and many other academic related matters: <https://www.uwo.ca/sci/counselling/>

Please contact the course instructor if you require lecture or printed material in an alternate format or if any other arrangements can make this course more accessible to you. You may also wish to contact Student Accessibility Services (SAS) at 661-2147 if you have any questions regarding accommodations.

Western University is committed to a thriving campus as we deliver our courses in the mixed model of both virtual and face-to-face formats. We encourage you to check out the Digital Student Experience website to manage your academics and well-being: <https://www.uwo.ca/se/digital/>.

The policy on Accommodation for Students with Disabilities can be found here: https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic%20Accommodation_disabilities.pdf

The policy on Accommodation for Religious Holidays can be found here: http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_religious.pdf

Learning-skills counsellors at the Student Development Centre (<http://www.sdc.uwo.ca>) are ready to help you improve your learning skills. They offer presentations on strategies for improving time management, multiple-choice exam preparation/writing, textbook reading, and more. Individual support is offered throughout the Fall/Winter terms in the drop-in Learning Help Centre, and year-round through individual counselling.

Students who are in emotional/mental distress should refer to Mental Health@Western (http://www.health.uwo.ca/mental_health) for a complete list of options about how to obtain help.

Additional student-run support services are offered by the USC, <http://westernusc.ca/services>.