

CS 4417/9647/9917b: Unstructured Data Course

The University of Western Ontario

Course Outline – Spring 2020

1. General Course Information

Course Information

CS 4417b: Unstructured Data

Tuesday: 4:30-5:30, SH-2355; Thursday 2:30-4:30, SH-2355.

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 – MC363
- Hanan Lutfiyya – hlutfiyy at uwo.ca – MC 355

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

3. Course Description

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://pdfs.semanticscholar.org/9d64/eaf01183ccbf3f2921a00ba3388c817bd72b.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:

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

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 examples are below:

- **Midterm exam:** February 13th 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 and assignments 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	January 31	Medium	14%
2	March 6	High	18%
3	April 3	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 midnight (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.

6. Accommodation and Accessibility

Self-Reporting Accommodation (SRA)

Any student who uses an SRA on an assignment will be given a 48 hour extension.

Any student who uses an SRA on the midterm will have the weight of the final 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.*

Note that if documentation (medical or otherwise) is required, it can only be collected by the student's Dean's Office/Academic Counselling unit.

If you are unable to meet a course requirement due to illness or other serious circumstances, you must seek approval for the absence as soon as possible. Approval can be granted either through a self-reporting of absence or via the Dean's Office/Academic Counselling unit of your Home Faculty. If you are a Science student, the Academic Counselling Office of the Faculty of Science is located in NCB 280, and can be contacted at scibmsac@uwo.ca.

For further information, please consult the university's policy on academic consideration for student absences:
https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Consideration_for_absences.pdf

If a student has received academic accommodation, missed components are excused so that the course is reweighted.

For accommodation for work worth 10% or less of the final grade, the student is referred to the student's Dean's Office/Academic Counselling unit.

If you miss the Final Exam, please contact your faculty's Academic Counselling Office as soon as you are able to do so. They will assess your eligibility to write the Special Exam (the name given by the university to a makeup Final Exam).

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).

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.

No electronic devices will be permitted on tests and exams.

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 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.

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>.