

CS 4417/9117/9647 - Unstructured Data Course Outline

1. Course Information

Unstructured Data

CS4417 (Computer Science Undergraduate)

CS9117 (Master of Data Analytics)

CS9647 (Computer Science Graduate)

Time/Place

NCB-113 (North Campus Building)

Tuesday 3:30pm to 4:20pm

Thursday 2:30pm to 4:20pm

List of Prerequisites

CS3319 A/B (Introduction to Databases)

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

Instructors	OWL Messaging	Email	Office	Phone	Office Hours
Dr. Daniel Lizotte (Course Coordinator)	“Instructor”	dlizotte@uwo.ca	MC363 WCPHFM4103	x86644	TBA
Chris Steward (Teaching Assistant)	“Secondary Instructor”	cstew2@uwo.ca			TBA
Juncheng Yin (Teaching Assistant)	“Secondary Instructor”	jyin97@uwo.ca			TBA

Communication must be directed as follows:

To manage the large volume of communication between students and the instructional team, **we will use OWL forums and messaging** as per the policy set out below. See Section 4 below for more information about OWL. *Communication sent via an inappropriate channel (e.g. an assignment clarification question sent by e-mail) will be ignored.*

- Questions about **course content and process that are relevant to the whole class** (e.g., specific and general assignment clarifications and questions, questions about midterm content) **must be directed to the OWL forums**. This way, others can benefit from the answers. Conversely, **students must check the forums prior to messaging**. Questions may be posted anonymously if a student wishes – *this hides their name from their classmates but not from the instructional team*. **Do not post any code or assignment answers in the OWL forums**.
- Questions about any **individual concerns**, for example about marks/marking, may be **sent by OWL message** to members of the instruction team:
 - Sending a message to the 'Secondary Instructor Role' contacts the TAs
 - **Questions about marks/marking should be directed toward TAs.**
 - Sending a message to the 'Instructor Role' contacts the professor.
 - Sending a message to all of us helps us to respond faster and know what's going on.
- **Course-related communication not sent via OWL messaging may be ignored**. You can set up OWL messaging to be relayed to your UWO e-mail if you prefer.

Messages from the instructional team will be sent by OWL announcement. You can set this to forward to your e-mail if you like; announcements will also be archived on the OWL site.

Office hours timing, personnel, and format will be conveyed through OWL as the term progresses.

3. Course Syllabus, Schedule, Delivery Mode

The objective of this course is to introduce students to techniques for the management, representation, 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 sources and management solutions with structured ones
- Recognize when unstructured data sources can support a data analytics solution
- Explain the role of representation learning in unstructured data analysis and provide examples
- Estimate the computation and storage needs for analyzing and storing a given large unstructured data source
- Use 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

Lectures

NCB-113 (North Campus Building)
 Tuesday 3:30pm to 4:20pm
 Thursday 2:30pm to 4:20pm

Office Hours

TBA (announced via OWL) based on Assignment, Midterm, and Exam needs
 Potentially a mix of in-person and Zoom

Contingency plan for an in-person class pivoting to 100% online learning

In the event of a COVID-19 resurgence during the course that necessitates the course delivery moving away from face-to-face interaction, affected course content will be delivered entirely online, either synchronously (i.e., at the times indicated in the timetable) or asynchronously (e.g., posted on OWL for students to view at their convenience). The grading scheme will **not** change. Any remaining assessments will also be conducted online as determined by the course instructor.

4. Course Materials

Presentation Materials

Materials from all lectures will be posted on OWL after class. They will not be made available prior to class.

Required Readings/Videos

Any required readings or videos will be posted on OWL at least one week in advance.

Suggested Textbooks

The following textbooks are recommended as references and to provide different/complementary explanations of some of the topics we will cover.

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

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

Stephan Butcher, 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; such references will be posted on OWL.

Students are responsible for checking the course OWL site (<http://owl.uwo.ca>) on a regular basis for news and updates. This is the primary method by which information will be disseminated to all students in the class.

All course material will be posted to OWL: <http://owl.uwo.ca>.

If students need assistance with the course OWL site, they can seek support on the OWL Help page. Alternatively, they can contact the Western Technology Services Helpdesk. They can be contacted by phone at 519-661-3800 or ext. 83800.

Technical Requirements

Students must have access to a computer on which they can install software and write code. Students are expected to manage their own software installations necessary to complete the coursework.

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 Final Exam is cumulative but with a focus on material covered after the Midterm Exam.

The following rules are designed to ensure that students meet the following 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 16th 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 with the following breakdown:

Assignment	Due Date	Expected Load	Weight (within assignment grade)
1	17 February	Medium	30%
2	19 March	High	35%
3	7 April	High	35%

- 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) on the due date. Late assignments may be handed in, but with a penalty of **5% of the assignment mark per day**. Saturday and Sunday together count as one day, so an assignment due on a Friday and handed in two days later on Sunday night at 11:54pm would have a 5% penalty. Assignments more than 5 days late will not be accepted. (E.g., an assignment due Friday will not be accepted after 11:55pm on the following Thursday.)
- 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 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 and on OWL. These will be due at 11:55pm on April 10th. **Late submissions will not be accepted.**

6. Student Absences

If you are unable to meet a course requirement due to illness or other serious circumstances, please follow the procedures below.

For work totalling 10% or more of the final course grade [*note: all of the assessments in this course meet this threshold*] you must provide valid medical or supporting documentation to the Academic Counselling Office of your Faculty of Registration as soon as possible. For further information, please consult the University's medical illness policy at

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

The Student Medical Certificate is available at

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

A missed midterm will have its weight transferred to the final exam.

A missed assignment will have its weight transferred proportionally to the other assignments.

Absences from Final Examinations

If you miss the Final Exam, please contact the Academic Counselling office of your Faculty of Registration as soon as you are able to do so. They will assess your eligibility to write the Special Examination (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" (e.g., more than 2 exams in 23-hour period, more than 3 exams in a 47-hour period).

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 the Academic Calendar for details (under [Special Examinations](#)).

Note: missed work can *only* be excused through one of the mechanisms above.

6. Accommodation and Accessibility

Religious Accommodation

When a course requirement conflicts with a religious holiday that requires an absence from the University or prohibits certain activities, students should request accommodation for their absence in writing at least two weeks prior to the holiday to the course instructor and/or the Academic Counselling office of their Faculty of Registration. Please consult University's list of recognized religious holidays (updated annually) at

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

Accommodation Policies

Students with disabilities are encouraged to contact Accessible Education, which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The policy on Academic Accommodation for Students with Disabilities can be found at:

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

7. Academic Policies

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

In accordance with policy,

https://www.uwo.ca/univsec/pdf/policies_procedures/section1/mapp113.pdf,

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 their official university address is attended to in a timely manner.

Electronic devices are not 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 the following Web site:

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>).

8. Support Services

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

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

Western is committed to reducing incidents of gender-based and sexual violence and providing compassionate support to anyone who has gone through these traumatic events. If you have experienced sexual or gender-based violence (either recently or in the past), you will find information about support services for survivors, including emergency contacts at

https://www.uwo.ca/health/student_support/survivor_support/get-help.html.

To connect with a case manager or set up an appointment, please contact support@uwo.ca.

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 Accessible Education at

http://academicsupport.uwo.ca/accessible_education/index.html

if you have any questions regarding accommodations.

Learning-skills counsellors at the Student Development Centre (<https://learning.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.

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/>.

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