CS 4417B/9117/9647 - Unstructured Data
Winter 2025 Term
Course Outline

1. Course Information

Course Information

Unstructured Data
CS4417 (Computer Science Undergraduate)
CS9117 (Master of Data Analytics)
CS9647 (Computer Science Graduate)

Time/Place

████████████████████████
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's Designate (Department/Program Counsellors and Science Academic Advisors) 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	Email	Office	Phone	Office Hours
Dr. Marwa Elsayed (Course Coordinator)	marwa.elsayed@uwo.ca	██████	██████	TBA
TBA Teaching Assistants				

Students must use their Western (@uwo.ca) email addresses when contacting the instructional team.

Communication must be directed as follows:

To manage the large volume of communication between students and the instructional team, **we will use OWL Brightspace forums and messages** 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 Brightspace 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 TAs
    - **Questions about marks/marking should be directed toward TAs.**
  - Sending a message to the 'Instructor' 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/discussions may be ignored.** You can set up OWL notifications 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.
- Recognize the process, applications, and key technologies of vector search.
- Describe embeddings and the LLM APIs used for embeddings.
- Build a search engine by using advanced Vector Search APIs.
- 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

███████████████

Tuesday 3:30pm to 4:20pm
Thursday 2:30pm to 4:20pm

Office Hours
TBA (announced via OWL Brightspace) based on Assignment, Midterm, and Exam needs
Potentially a mix of in-person and Microsoft Teams.

Important Academic Sessional Dates:
Classes begin: January 6, 2025
Spring Reading Week: February 15 – 23, 2025
Classes end: April 4, 2025
Exam period: April 7 – 30, 2025

Contingency plan for an in-person class pivoting to 100% online learning
In the event of a university declared emergency 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 before class will be posted on OWL 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. There is no cost for these textbooks as they can be accessible online.

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.wordpress.com/wp-content/uploads/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 inside class presentations.

All course material will be posted to OWL: https://westernu.brightspace.com/

Students are responsible for checking the course OWL site (https://westernu.brightspace.com/) regularly for news and updates. This is the primary method by which information will be disseminated to all students in the class.

If students need assistance with the course OWL site, they can seek support on the OWL Brightspace 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

Grading Scheme and Assessment Dates
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.

Exams
The dates for the exams are below:
- **Midterm exam:** February 13th online
- **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 Consideration 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 Brightspace. 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	7 February	Medium	30%
2	14 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.
- 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 7th. **Late submissions will not be accepted.**

General information about missed coursework

This section is only applicable to 4417 students, and hence all graduate students taking the course are required to meet stated deadlines. Students must familiarize themselves with the *University Policy on Academic Consideration – Undergraduate Students in First Entry Programs* posted on the Academic Calendar:
https://www.uwo.ca/univsec/pdf/academic_policies/appeals/academic_consideration_Sep24.pdf,

This policy does not apply to requests for Academic Consideration submitted for **attempted or completed work**, whether online or in person.

The policy also does not apply to students experiencing longer-term impacts on their academic responsibilities. These students should consult Accessible Education.

For procedures on how to submit Academic Consideration requests, please see the information posted on the Office of the Registrar's webpage:
https://registrar.uwo.ca/academics/academic_considerations/
All requests for Academic Consideration must be made within 48 hours after the assessment date or submission deadline.

All Academic Consideration requests normally must include supporting documentation; however, recognizing that formal documentation may not be available in some extenuating circumstances, the policy allows students to make one Academic Consideration request without supporting documentation in a course. The following assessments are excluded from this, and therefore always require formal supporting documentation:
- Midterm Exam

However, given the built-in flexibility for the assignments, and the above protection, all course components require documentation to support academic consideration requests.

Evaluation Scheme for Missed Assessments
- A missed midterm will have its weight transferred to the final exam when academic consideration has been granted.
- A missed assignment will have its weight transferred proportionally to the other assignments when academic consideration has been granted.

Absences from Final Examinations
When a student misses the Final Exam and their Academic Consideration has been granted, they will be allowed to write the Special Examination (the name given by the University to a makeup Final Exam). See the Academic Calendar for details (under Special Examinations), especially for those who miss multiple final exams within one examination period.

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

Essential Learning Requirements
Even when Academic Considerations are granted for missed coursework, the following are deemed essential to earn a passing grade.

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.

Coursework with Assessment Flexibility
By policy, instructors may deny Academic Consideration requests for the following assessments with built-in flexibility:

Flexible Completion
- **Assignments.** Students are expected to submit each of the three assignments by 11:55pm (electronically) on the deadline listed. Should extenuating circumstances arise, students do not need to request Academic Consideration and they are permitted to submit their assignment up to (24 hours) past the deadline without a late penalty. Should students submit their assessment beyond (24 hours) past the deadline, **a late penalty of 5% per day will be applied**. Academic Consideration requests may be granted only for extenuating circumstances that started before the deadline and lasted longer than the No-Late-Penalty Period (24 hours). **Assignments more than 5 days late will not be accepted even with Academic Consideration granted.** (e.g., an assignment due Friday will not be accepted after 11:55pm on the following Thursday.)

6. Additional Statements

Religious Accommodation
When conflicts with a religious holiday that requires an absence from the University or prohibits certain activities, students should request an accommodation for their absence in writing to the course instructor and/or the Academic Advising office of their Faculty of Registration. This notice should be made as early as possible but not later than two weeks prior to the writing or the examination (or one week prior to the writing of the test).

Please visit the Diversity Calendars posted on our university's EDID website for the recognized religious holidays: https://www.edi.uwo.ca.

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.

Academic Policies

The website for Registrar Services is https://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 will not 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 the following Web site: https://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.

Programming assignments may be subject to submission for similarity review and detection by Gradescope, which uses Measure Of Software Similarity (MOSS). MOSS is a tool which detects similar logic and flow in programs. https://theory.stanford.edu/~aiken/moss/

The iClicker mobile app may be used during class to foster interactive learning and enhance student learning outcomes. It may be used to conduct polls, facilitate discussions, and collect instant feedback on student comprehension. If so, students are required to authenticate using their Western email account; hence, it cannot be used by anyone other than the student.

Tests and examinations in this course will be conducted using a remote proctoring service if it is required to pivot to an online learning delivery model. By taking this course, you are consenting to the use of this software and acknowledge that you will be required to provide **personal information** (including some biometric data), and the session will be **recorded**. Completion of this course will require you to have a reliable internet connection and a device that meets the technical requirements for this service. More information about this remote proctoring service, including technical requirements, is available on Western's Remote Proctoring website at: https://remoteproctoring.uwo.ca.

Using GenAI tools like ChatGPT can be beneficial when used ethically and transparently. Students should limit their use to brainstorming ideas, clarifying concepts, or gaining a deeper understanding of the material. However, students must remain aware that AI-generated content can contain misinformation or inaccurate explanations. These tools must not be used for answering course assignments or completing deliverables, as doing so would be considered a compromise of academic integrity. If a student uses GenAI tools they must write a paragraph stating their exact use and submit it along with their submission. By using AI thoughtfully within these guidelines, students can enhance their learning while developing essential skills such as critical thinking, problem-solving, and research, contributing to their academic success in this course.

In short, all work submitted in this course should be your OWN work. If you are in doubt about whether something is a breach of academic dishonesty, please ask before doing it.

Reasonable
• Working together on practice exercises, sharing solutions, posting solutions to practice problems and/or old exams is permitted.

• Discussing assignments verbally (or via the forums) to better understand or to plan a solution is permitted.

• Incorporating snippets of code that you find on-line to help you solve the problem as long as the code is NOT the solution to the problem. The URL containing the source of the code that you use must be cited in the comments. Failure to do so will be considered a breach of academic integrity.

Unreasonable
• Viewing a classmate's solution.

• Providing a classmate with a solution.

• Failing to cite code sources that you have used from the web.

• Submitting a copy (even with minor modifications) of someone else's code. (Beware that people are as unique in coding as they are in writing -- it is quite easy to determine if two assignments are similar).

• Posting your solution on a public (or private) web site for others to view.

• Soliciting solutions to the problem on on-line forums or purchasing solutions from online sources.

• Submitting a combined solution with a classmate.

Support Services
Please visit the Science & Basic Medical Sciences Academic Advising 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 Learning Development and Success (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/.