

Natural Language Processing

Computer Science

2024 Winter

**Instructor Information**

Instructor: Dr. Amin Safaei

Office Location: AT5024

Telephone: TBA

E-mail: asafaei@lakeheadu.ca

Office Hours: **Thursday and Friday** at 12:00 pm to 1:00 pm (by email appointment)

**Teaching Assistant (TA) Information**:

TBA

**Course Identification**

Course Number: COMP-5014-WA

Course Name: Natural Language Processing

Course Location: **RB-1022**

Class Times: **Friday** 02:30PM-05:30PM

Prerequisites: ---

**Course Description/Overview**

Natural Language Processing (NLP) is a rapidly developing field with broad applicability throughout the hard sciences, social sciences, and the humanities. The ability to harness, employ and analyze linguistic and textual data effectively is a highly desirable skill for academic work, in government, and throughout the private sector.

This course is intended as a theoretical and methodological introduction to a the most widely used and effective current techniques, strategies and toolkits for natural language processing, with a primary focus on those available in the Python programming language.

**Course Outline**

* Introduction to Natural Language Processing (NLP)
* Processing text for NLP
* Implementing Sentiment Analysis
* Introducing Information Extraction
* Introduction to topic modeling
* Working with languages

**Course Resources**

Course Website(s)

* myCourseLink

Required Course Text(s)

* There is no required textbook for this course. The instructor will provide the students with class notes, slides, and other study materials.

Additional text

* Jurafsky, David, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.

Software environment

* All the software used in this course is either open source or available in the computer labs.
* See D2L for the details.

**Take-Home Assignments**

* There will be four assignments this term that **must submit through d2l (mycourselink).**
* **Sending assignment to instructor email or TAs will not evaluate and receive 0.**
* **Late submission is not accepted.**

**Assignments and Evaluations**

|  |  |
| --- | --- |
| * Assignment: (5+5+5+5) | 20% |
| * Mid-term exam (Closed book) | 20% |
| * Final exam (Closed book) | 30% |
| * Project | 20% |
| * Attendance and Pop Quizzes | 10% |

**Attendance**

* Regular attendance at lectures as well as active participation in classroom discussions is an important factor for student success. The successful student must develop good note-taking skills and be willing to ask questions. The attendance of lectures with the submission of pop quizzes will contribute 10% to the final grade.
* If you are not able to attend the session, then you must inform me ahead of time ( at least by one working day). By not doing so, will result in a penalty of 2% in your final grades.

**Course Schedule/Outline**

|  |  |  |
| --- | --- | --- |
| **Week** | **Description** | **Date** |
| **Wednesday** |
| 1 | Introduction to NLP  Common NLP tasks | JAN 12, 2024 |
| 2 | Walkthrough of an NLP problem  Evolution of NLP architectures | JAN 19, 2024 |
| 3 | Text processing overview  Getting text | JAN 26, 2024 |
| 4 | Text preprocessing  Vectorizing text | FEB 2, 2024 |
| 5 | Advanced processing  Storing and visualizing unstructured data | FEB 9, 2024 |
| 6 | Identifying the steps for text processing  Examining the algorithms for sentiment analysis | FEB 16, 2024 |
| 7 | Study Week – No Classes | FEB 23, 2024 |
| 8 | Midterm Exam | MAR 1, 2024 |
| 9 | Types of information extraction  Implementing information extraction | MAR 8, 2024 |
| 10 | Identifying the approach  Implementing topic modeling | MAR 15, 2024 |
| 11 | Detecting and translating languages  Transcribing and vocalizing text | MAR 22, 2024 |
| 12 | Project Presentation 1 | MAR 29, 2024 |
| 13 | Project Presentation 2 | APR 5, 2024 |

**Course Policies**

* behavioral standards (this could be a reference to the [Student Code of Conduct - Academic Integrity](https://www.lakeheadu.ca/students/student-life/student-conduct/academic-integrity/node/51239))
* attendance,
* safety regulations (e.g. completion of WHMIS training for lab work, PPE, etc.)
* netiquette (e.g. set guidelines about how students should communicate with you including hours of availability, reasonable response time, subject headings, salutation, etc.),

**Academic Integrity Statement (**Exam/Assignment Integrity)**:**

I understand and agree that:

(1) Unless otherwise allowed by the course instructor, I must complete the assignments in this course without the assistance of anyone else.

(2) Unless otherwise allowed by the course instructor, I must not access any sources or materials (in print, online, or in any other way) to complete any course exam.

I further understand and agree that, if I violate either of these two rules, or if I provide any false or misleading information about my completion of course assignments or exams, I may be prosecuted under the Lakehead University Student Code of Conduct – Academic Integrity, which requires students to act ethically and with integrity in academic matters and to demonstrate behaviors that support the University’s academic values.

**Copyright**

Students should be aware that all instructional, reference, and administrative materials prepared for this course are protected in their entirety by copyright. Students are expected to comply with this copyright by only accessing and using the course materials for personal educational use related to the course, and that the materials cannot be shared in any way, without the written authorization of the course instructor. If this copyright is infringed in anyway, students may be prosecuted under the Lakehead University Student Code of Conduct – Academic Integrity, which requires students to act ethically and with integrity in academic matters and to demonstrate behaviors that support the University’s academic values.

**Copyright Compliance (D2L):**

I understand and agree that all instructional, reference, and administrative materials to which I am given access in this course (the "course materials"), whether they consist of text, still or kinetic images, or sound, whether they are in digital or hard copy formats, and in whatever media they are offered, are protected in their entirety by copyright, and that to comply with this copyright and the law

(1) I may access and download the course materials only for my own personal and non-commercial use for this course; and

(2) I am not permitted to download, copy, or store (in any medium) any text, image, or sound component of the course materials for any other purpose whatsoever, or to forward or share, transmit, broadcast, show, post or play in public, adapt, or change in any way any text, image, or sound component of the course materials, except as expressly authorized, and only to the extent authorized, in writing, by the course instructor."

I further understand and agree that, if I infringe the copyright of the course materials in any way, I may be prosecuted under the Lakehead University Student Code of Conduct – Academic Integrity, which requires students to act ethically and with integrity in academic matters and to demonstrate behaviours that support the University’s academic values.

**Regulations**

It is the responsibility of each student registered at Lakehead University to be familiar with, and comply with all the terms, requirements, regulations, policies and conditions in the Lakehead University [Academic Calendar](http://navigator.lakeheadu.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=25&chapterid=6999&loaduseredits=False). This includes, but is not limited to, Academic Program Requirements, Academic Schedule of Dates, University and Faculty/School Policies and Regulations and the Fees and Refund Policies and Schedules (Lakehead University Regulations webpage, 2020-21).

**Academic Integrity**

A breach of Academic Integrity is a serious offence. The principle of Academic Integrity, particularly of doing one’s own work, documenting properly (including use of quotation marks, appropriate paraphrasing and referencing/citation), collaborating appropriately, and avoiding misrepresentation, is a core principle in university study. Students should view the [Student Code of Conduct - Academic Integrity](https://www.lakeheadu.ca/students/student-life/student-conduct/academic-integrity/node/51239) for a full description of academic offences, procedures when Academic Integrity breaches are suspected and sanctions for breaches of Academic Integrity.

**Supports for Students** – there are many resources available to support students. These include but are not limited to:

* [Health and Wellness](https://www.lakeheadu.ca/current-students/student-health-and-counselling-centre)
* [Student Success Centre](https://www.lakeheadu.ca/current-students/student-success-centre)
* [Student Accessibility Centre](https://www.lakeheadu.ca/current-students/student-services/accessibility/)
* [Library](https://library.lakeheadu.ca/)
* [Lakehead International](https://www.lakeheadu.ca/international)
* [Indigenous Initiatives](https://www.lakeheadu.ca/indigenous)

Lakehead University is committed to achieving full accessibility for persons with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities and/or medical conditions to ensure they have an equitable opportunity to participate in all of their academic activities. If you are a student with a disability and think you may need accommodations, you are strongly encouraged to contact Student Accessibility Services (SAS) and register as early as possible. For more information, please contact [Student Accessibility Services](http://studentaccessibility.lakeheadu.ca/)  (SC0003, 343-8047 or [sas@lakeheadu.ca](mailto:sas@lakeheadu.ca)