CSCI 4152 — Natural Language Processing
CSCI 6509 — Advanced Topics in Natural Language Processing
Course Syllabus

Instructor Information

Term:	Fall 2023	
Instructor:	Dr. Vlado Keselj	
E-mail:	`vlado@cs.dal.ca`	
Other e-mails:	`vlado@dnlp.ca` `vkeselj@dal.ca` `Vlado.Keselj@dal.ca`	
	(`@dnlp.ca` email is managed by a server in the Faculty of Computer Science, Dalhousie University campus)	
TA:	TBA	
Lectures:	Tue-Thu 16:05–17:25	Rowe 1011
Labs CSCI 4152:	Tue 17:35–18:55	Goldberg CS 134
Labs CSCI 6509:	Wed 17:35–18:55	Goldberg CS 143
Website:	`http://web.cs.dal.ca/~vlado/csci6509`	
	Brightspace site is also available to registered students.	
E-mail List:	`nlp-course@lists.dnlp.ca`	

Important Dates

Please check the course calendar on the web site for more details.

1. Tue Sep 5, 2023: Term starts, Lectures begin
2. Tue Sep 19, 2023: Last day to add/drop classes
3. Mon Oct 2, 2023: National Day for Truth and Reconciliation, University closed
4. Wed Oct 4, 2023: Last day to drop without "W", change audit to credit or vv.
5. Mon Oct 9, 2023: Thanksgiving Day, University closed
6. Thu Nov 2, 2023: Last day to drop with "W"
7. Nov 13–17, 2023: Fall Study Break, no classes, University open except on Monday
8. Mon Nov 13, 2023: in lieu of Remembrance Day, University closed
9. Tue Dec 5, 2023: Monday schedule on Tuesday
10. Wed Dec 6, 2023: Classes end, Reports due, Monday schedule on Wednesday
11. Dec 8–19, 2023: Exam period

Course Description

Natural Language Processing (NLP) is an area of Computer Science, and sub-area of Artificial Intelligence, concerned with the problem of automatically processing natural languages in written and spoken form. The course covers the most common methodology in analyzing and generating natural languages, such as English, French, and other naturally developed languages. This course introduces fundamental concepts and principles used in NLP with emphasis on two approaches to NLP: statistical and unification-based. Some applications are discussed, such as the problems of text classification, information extraction, and question answering.

Course Rationale

The course provides an introduction into Natural Language Processing (NLP), which is an important area of Computer Science and Artificial Intelligence. The lecture material covers the main NLP methodology and applications, and the course project material and requirements are aimed at experiental learning in NLP implementation for undergraduate students, and NLP research for graduate and optionally undergraduate students.

Tentative List of Topics

1. Introduction to NLP
2. Stream-based Text Processing

   - Automata and Regular Expressions
   - Text Processing using Perl
   - Elements of Morphology
   - Elements of Information Retrieval
   - Character N-gram Based Text Mining
   - Text Classification

3. Probabilistic Approach to NLP

   - Probabilistic Modeling
   - Text Processing using Python
   - Naïve Bayes Model
   - N-grams Model and Language Modeling
   - Hidden Markov Model and Bayesian Networks
   - Sum-product Algorithms
   - Neural Network Models

4. Syntactic Processing

   - Phrase Structure and NL Syntax
   - Chart Parsing and CYK Algorithm
   - Probabilistic Context-Free Grammars

5. Unification-based NLP and Semantics

   - First-order Predicate Logic and Resolution
   - Text Processing using Prolog
   - Classical and Feature-structure Unification
   - Unification-based Grammars

6. Student Presentations

Learning Outcomes

- Describe the area of Natural Language Processing.
- Describe the levels of NLP.
- Select appropriate levels of NLP based on an application task.
- Write an NFA or DFA for a text processing task.
- Translate NFA into DFA.
- Use regular expressions to solve basic text preprocessing tasks.
- Use Perl and Python in text processing.
- Describe use of Prolog and DCFG grammar in NL parsing.
- Apply basic character n-gram techniques in text classification.
- Describe probabilistic modeling and associated computational tasks.
- Describe, use, and evaluate the Naïve Bayes, Hidden Markov, and Bayesian Network models.
- Use sum-product algorithms to solve inference tasks in a structured probabilistic model.
- Describe, use, and evaluate chart parsing using the CYK and similar algorithms.
- Describe and apply Probabilistic Context-Free Model.
- Describe the concept of unification-based grammars.
- Describe classical and feature-structure unification.
- Propose, develop, evaluate and report on a software solution in the NLP area.

Class Format and Course Communication

- Content will be delivered using lectures and labs in-person. It is possible that some classes may be on-line, or in a hybrid mode, using a platform such as MS Teams.

- Course announcements will be posted to the course email list, which comprises the instructor's, the TA's, and the students' email accounts. Based on the Banner registration information, the Dalhousie email addresses of the students will be initially added to the list. The students may request additional addresses to be added. Important course announcements may be posted on the list, so if there are any issues with receiving emails from the list, the students should get in touch with the instructor to have the issue resolved. It is the student's responsibility to check their e-mail on a daily basis.
- Most course deliverables will be in the electronic form. The submissions may include submission by email, submission via a web site, submission via Dal FCS GitLab server, and possibly other forms of submission. Some submissions may be on paper, and the final exam will be a 2-hour exam in person in December.

Evaluation Criteria

Overview:	32%	Assignments
	32%	Final exam
	10%	Class Presentation and Participation
	26%	Project Deliverables (P0, P1, Report)

1. Assignments
   - Tentatively 5 assignments (possibly more smaller assignments)
   - **Late assignments are not accepted.** If you think that you cannot finish assignment in time, it must be resolved with the instructor before the deadline. **The SDA forms** (Student Declaration of Absence) can be used for assignment deadlines, but **not if the deadline was already postponed.**
   - No collaboration is permitted on the assignments.
   - Assignments will be submitted electronically; typically via the course web site. Detailed instructions will be included in the assignment.
   - Assignments may include in-class assignments (i.e., tests) and computer practicums

2. Final Exam
   - Scheduled by the university.
   - Will cover all lecture and lab material in the course (student projects not included).

3. Project presentation and class participation
   - One short presentation during the last few weeks of the term
   - Includes class participation and attendance

4. Project Deliverables
   - The final report and intermediate updates

Final Exam Requirements

- Photo ID is required.
- Closed book; a simple calculator is allowed.

Required Texts and Resources

The course lecture slides and notes, and lab notes are posted on the course website during the term. They are generally posted a few days or shortly before a lecture or lab, and updated after the lecture or lab.

The required textbook is:

- *Speech and Language Processing,* by Daniel Jurafsky and James Martin, 2013.

A few more recommended books are listed on the course web site.

The rest of the syllabus are a reminder of Faculty of Computer Science and Dalhousie policies that are applicable to all CS courses.

Reminder of Some Policies and Other Information
Related to CS Courses at the Dalhousie

Student Declaration of Absence

The Student Declaration of Absence forms information: `https://www.dal.ca/campus_life/safety-respect/student-rights-and-responsibilities/academic-policies/student-absence.html`

Academic Standards

Failure to properly attribute sources in your work will be treated as an academic standards issue and points may be deducted for not following citation requirements. For example, forgetting to quote text taken from other sources, failure to include in-text citations, or a failure to include required information in the citations or references. Please see the resources on proper citation provided by the Dalhousie Writing Center (`https://dal.ca.libguides.com/c.php?g=257176&p=5001261`).

Please note that if it appears that the error was made with intent to claim other people's work as your own such as a lack of both citations and references, an allegation of plagiarism will be submitted to the Faculty Academic Integrity Officer, which could result in consequences such as a course failure.

Responsible Computing Policy

Usage of all computing resources in the Faculty of Computer Science must be within the Dalhousie Acceptable Use Policies (`https://www.dal.ca/dept/university_secretariat/policies/information-management-and-technology/acceptable-use-policy-.html`) and the Faculty of Computer Science Responsible Computing Policy. (`https://www.dal.ca/content/dam/dalhousie/pdf/faculty/computerscience/policies-procedures/fcs_policy_local.pdf`)

Use of Plagiarism Detection Software

All submitted assignment may be passed through a plagiarism detection software, such as the Moss Software Similarity Detection System (`https://theory.stanford.edu/~aiken/moss/`), or similar systems, after simple automated anonymization. If a student does not wish to have their assignments passed through plagiarism detection software, they should contact the instructor for an alternative. Please note, that code not passed through plagiarism detection software will necessarily receive closer scrutiny. `https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf`

Use of Artificial Intelligence Tools

You may use AI-driven tools to assist you in learning but remember that your objective is to understand, achieve, and apply the course competencies and outcomes. While you may use tools for learning, the assessments in this course will disallow the use of AI-driven tools to assert that you have attained course learning outcomes. This is because a graduate must be able to analyze, assess and produce work unassisted by AI technology. Where tools are allowed: you must acknowledge all tools used to assist you. If applicable, you must provide links to chat logs. Using AI-driven tools where pro-hibited constitutes an academic offense.

Submitting assignments, lecture notes, and similar material provided by the instructor in the course to the third-party sites, such as ChatGPT, CourseHero, Chegg, and similar is not allowed.

Culture of Respect

Every person has a right to respect and safety. We believe inclusiveness is fundamental to education and learning. Misogyny and other disrespectful behaviour in our classrooms, on our campus, on social media, and in our community is unacceptable. As a community, we must stand for equality and hold ourselves to a higher standard.

What we all need to do [1]:

1. **Be Ready to Act:** This starts with promising yourself to speak up to help prevent it from happening again. Whatever it takes, summon your courage to address the issue. Try to approach the issue with open-ended questions like "Why did you say that?" or "How did you develop that belief?"
2. **Identify the Behaviour:** Use reflective listening and avoid labeling, name-calling, or assigning blame to the person. Focus the conversation on the behaviour, not on the person. For example, "The comment you just made sounded racist, is that what you intended?" is a better approach than "You're a racist if you make comments like that."
3. **Appeal to Principles:** This can work well if the person is known to you, like a friend, sibling, or co-worker. For example, "I have always thought of you as a fair-minded person, so it shocks me when I hear you say something like that."
4. **Set Limits:** You cannot control another person's actions, but you can control what happens in your space. Do not be afraid to ask someone "Please do not tell racist jokes in my presence anymore" or state "This classroom is not a place where I allow homophobia to occur." After you have set that expectation, make sure you consistently maintain it.
5. **Find or be an Ally:** Seek out like-minded people that support your views, and help support others in their challenges. Leading by example can be a powerful way to inspire others to do the same.
6. **Be Vigilant:** Change can happen slowly, but do not let this deter you. Stay prepared, keep speaking up, and do not let yourself be silenced.

University Statements

This course is governed by the academic rules and regulations set forth in the University Calendar and the Senate. `https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx`

Territorial Acknowledgement
> Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

Internationalization
> At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders." `https://www.dal.ca/about-dal/internationalization.html`

Academic Integrity
> At Dalhousie University, we are guided in all of our work by the values of academic integrity: honesty, trust, fairness, responsibility and respect. As a student, you are required to demonstrate these values in all of the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity.
> (read more: `http://www.dal.ca/dept/university_secretariat/academic-integrity.html`)

Accessibility
> The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online

---

[1]Source: Speak Up! ©2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full "Speak Up" document found at: `https://www.dal.ca/campus_life/safety-respect/student-rights-and-responsibilities/preventing-conflict/speak-up.html` Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University 902.494.4140 lyndsay.anderson@dal.ca www.dal.ca/think.

or in-person) that result in barriers to your inclusion please contact:
`https://www.dal.ca/campus_life/academic-support/accessibility.html`
for all courses offered by Dalhousie with the exceptionof Truro.

Conduct in the Classroom — Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

Diversity and Inclusion — Culture of Respect

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2). (read more:`http://www.dal.ca/cultureofrespect.html`)

Student Code of Conduct

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students dont follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal mannerperhaps through a restorative justice process. If an informal resolution cant be reached, or would be inappropriate, procedures exist for formal dispute resolution. (read more: `https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/Code%20of%20Student%20Conduct%20rev%20Sept%202021.pdf`)

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. (read more: `https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.html`)

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. (read more: `https://www.dal.ca/dept/university_secretariat/policies/academic/student-submission-of-assignments-and-use-of-originali html`)

Student Use of Course Materials

These course materials are designed for use as part of the CSCI courses at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading material to a commercial third party website) may lead to a violation of Copyright law.

Learning and Support Resources

Please see `https://www.dal.ca/campus_life/academic-support.html`