

UTCL : Natural Language Learning

Syllabus: LIN 386M, Natural Language Learning

Course Information

- Course: Natural Language Learning, LIN386M – 41245
- Semester: Spring 2010
- Course page: comp.ling.utexas.edu/courses/2010/spring/natural_language_learning
- Course times: Tuesday, Thursday 2:00–3:30pm
- Course location: MEZ 1.120 [<http://www.utexas.edu/maps/main/buildings/mez.html>].

Instructor Contact Information

- Katrin Erk [http://comp.ling.utexas.edu/people/katrin_erk]
- office hours: Mon 10am–12pm, Tue 3:30–4:30pm
- office: Calhoun 512 [<http://www.utexas.edu/maps/main/buildings/cal.html>]
- phone: 471–9020
- fax: 471–4340
- email: [katrin dot erk at mail dot utexas dot edu](mailto:katrin_dot_erk_at_mail_dot_utexas_dot_edu)

Lab information

To get an account on the computational linguistics lab machines, please contact our lab administrator, Joey Frazee, at

- [jfrazee at mail dot utexas dot edu](mailto:jfrazee@utexas.edu)

State that you are in this class, give a login name that you would like to have, and cc me on your email.

Prerequisites

Graduate standing. Computational Linguistics I (LIN 386.M), Natural Language Processing (CS388), Machine Learning (CS391L), or consent of instructors. Previous programming experience.

Syllabus and Text

This page serves as the syllabus for this course.

There is no official course text book. Selected readings will be made available for download or

copying.

Both Jurafsky and Martin, "Speech and language processing", and Manning and Schütze, "Foundations of statistical natural language processing", are likely to be useful resources.

Exams and Assignments

There will be no midterm or final exam. Instead, this course has a course project.

A tentative schedule for the entire semester is posted on the schedule page. Readings and exercises may change up one week in advance of their due dates.

To see your grades go to eGradebook [<https://utdirect.utexas.edu/diia/egb>].

Philosophy and Goal

Lexical information is one of the central bottlenecks in computational linguistics. Information retrieval applications need information on the semantic relatedness of words and documents. Sentiment analysis needs information about whether words can be used to express positive or negative opinions. Textual entailment (a meta-task encompassing problems from information retrieval, information extraction, summarization, and question answering) needs huge paraphrase resources, as well as deeper information about negation, polarity and factivity. In this seminar, we will discuss recent research papers on unsupervised and mildly supervised methods for learning lexical information from corpora.

Some specific goals of the course are to enable students to:

- gain an insight into different types of lexical information that can be learned from text, such as semantic classes, semantic and relational similarity, relations between concepts, and multi-word paraphrases
- gain exposure to important techniques in natural language processing, in particular vector space models, random walk models, and graph clustering
- understand how crosslingual information can be exploited in learning lexical information
- complete an advanced NLP project and write a report in the format of submissions to computational linguistics conferences

This course presents an opportunity for students to grapple directly with models used in automated natural language learning that underly practical applications while gaining an appreciation for the theoretical questions which they raise and which they can help us tackle. It will thus help prepare the student both for jobs in the industry and for doing original research in computational linguistics.

The course is designed to include key activities engaged in by computational linguistics researchers, including generation of ideas and programs, critical oral discussion of ideas, and written evaluation and presentation of ideas. This will help students make the transition to doing real research in the field. It is expected that this class will directly lead to concrete research opportunities for students.

Content Overview

This course provides an in-depth discussion of several important and current topics in automated natural language learning. One recurring question in this course will be how much lexical knowledge, and what kinds of lexical knowledge, can be learned from text. The course also takes up two themes from this year's SemEval [<http://semeval2.fbk.eu/>] workshop: cross-lingual approaches, and the interpretation of ambiguous phrases (such as noun compounds) through paraphrasing.

Topics include:

- learning semantic similarity through either vector space or random walk models
- learning (multi-word) paraphrases
- inferring relations between words or concepts
- unsupervised and weakly supervised methods for determining word meaning in context
- selectional preferences
- script learning
- learning facets of lexical meaning that are needed for deeper inferences
- exploiting crosslingual information in natural language learning

A detailed schedule for the course, with topics for each lecture, is available at http://comp.ling.utexas.edu/courses/2010/spring/natural_language_learning/schedule [http://comp.ling.utexas.edu/courses/2010/spring/natural_language_learning/schedule]

Course Requirements and Grading Policy

A note on course projects

Course projects can be done individually or in teams. If you decide to work in a team, you can choose to submit a common project proposal, project progress report, and project final report. In that case, those documents can be longer than given below, up to <number of team members> times <length given below>. In that case, indicate each team member's contributions in your texts. Alternatively, each team member can submit individual papers.

For course projects, you are encouraged to choose a task from this year's SemEval [<http://semeval2.fbk.eu/>]. You are also free to choose an independent topic.

Course requirements

Reading presentation (20%):

You will give a 20 minute presentation on one of the papers listed in the readings for the class, and lead the subsequent discussion about the paper. The grade will be based mainly on clarity of the presentation.

Project proposal (10%):

Midway through the semester, you will propose a topic for your final project. Topic ideas will be discussed in class prior to the project proposal draft due date. The proposal will be in written form and should be roughly 2-3 single-spaced pages and contain at least 5 references, done using LaTeX and the ACL submission style [http://comp.ling.utexas.edu/courses/2008/cl1_08/latex_for_proposal.tgz]. The draft will be evaluated primarily on written expression, coherence of argument, and originality of

research topic. Feedback will be given both on writing and content.

Project progress report (10%):

The progress report is mainly a revision and extension of the proposal. It should take into account all comments given on the proposal. Expect it to require significant rewriting, as opposed to just editing of the proposal. In addition, it should include an update on progress to date. It should be 4 pages in length and contain at least 8 references, using LaTeX and the ACL submission style [http://comp.ling.utexas.edu/courses/2008/cl1_08/latex_for_proposal.tgz]. It will be graded primarily on written expression and coherence of argument. Feedback will be given both on writing and content.

Project final report (40%):

The final report builds on the progress report and presents the project results and conclusions. It should be 8 pages in length and contain at least 10 references, using LaTeX and the ACL submission style [http://comp.ling.utexas.edu/courses/2008/cl1_08/latex_for_proposal.tgz]. The grade will be based on the final product (program, corpus, etc) and the written report.

Project presentation (10%):

Each student will give a 20 minute presentation on his or her project in the last week of class. The grade will be based mainly on clarity of the presentation.

Class participation (10%):

Students are expected to be present in class having completed the readings and participate actively in the discussions.

Grading policy

Final grades will use plus/minus grades.

Extension Policy

If you turn in your assignment late, expect points to be deducted. Extensions will be considered on a case-by-case basis.

In most cases they will not be granted.

In the event that an extension is granted, points will be deducted for lateness. By default, 5 points (out of 100) will be deducted for lateness, plus an additional 1 point for every 24-hour period beyond 2 that the assignment is late. For example, an assignment due at 12:30pm on Tuesday will have 5 points deducted if it is turned in late but before 12:30pm on Thursday. It will have 6 points deducted if it is turned in by 12:30pm Friday, etc.

The greater the advance notice of a need for an extension, the greater the likelihood of leniency.

Academic Dishonesty Policy

You are encouraged to discuss assignments with classmates. But all written work must be your own. Students caught cheating will automatically fail the course. If in doubt, ask the instructor.

Notice about students with disabilities

The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. Please contact the Division of Diversity and Community Engagement, Services for Students with Disabilities, 471-6259.

Notice about missed work due to religious holy days

A student who misses an examination, work assignment, or other project due to the observance of a religious holy day will be given an opportunity to complete the work missed within a reasonable time after the absence, provided that he or she has properly notified the instructor. It is the policy of the University of Texas at Austin that the student must notify the instructor at least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holy days that fall within the first two weeks of the semester, the notice should be given on the first day of the semester. The student will not be penalized for these excused absences, but the instructor may appropriately respond if the student fails to complete satisfactorily the missed assignment or examination within a reasonable time after the excused absence.