

University of Illinois at Chicago  
Spring 2006

## CS 421 — Natural Language Processing Course Syllabus

**Room:** 220 SH

**Time:** TuTh 9:30 – 10:45

**URL:** via Blackboard

### Staff

**Instructor:** Barbara Di Eugenio

**Office:** 916 SEO, 6-7566

**E-mail:** bdieugen@cs.uic.edu

**Office Hours:** Tue 2-3:30

**Teaching Assistant:** Rajen Subba

**Office:** 940 SEO

**E-mail:** rsubba@cs.uic.edu

**Office Hrs:** Th 2-3:30

### Course Objectives

The aim of this course is to introduce students to the field of Natural Language Processing (NLP), also called Computational Linguistics, or Natural Language Engineering. NLP studies the algorithms to enable the computer to interpret and produce *natural languages*, i.e. English, Russian, Mandarin Chinese, Italian, Turkish, etc. Applications of NLP include information retrieval from the web (GOOGLE explicitly mentions NLP in its recruitment ads), human computer interaction, and machine translation.

The course will provide students with the linguistic foundations that underlie NLP, introduce them to the algorithms used in the field, and provide practice in building components of NLP systems.

### Reading Materials

Required Textbook: Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, 2000

I will also distribute articles and other materials, some of them taken from the following books:

1. Christopher Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999
2. James Allen. *Natural Language Understanding* (2nd edition). The Benjamin/Cummings Publishing Company, 1995.

### Prerequisites

CS 301.

## Notes

- I use email a lot to communicate with the whole class. Please check your email frequently, especially around deadlines (homeworks and exams).
- The web page will contain all materials relevant to the class, syllabus, assignments etc. You can also see you own grades.
- When you have a question of general interest, rather than sending me or the TA mail, post your questions to the discussion board on the web site

## Tentative Schedule

Dates	Topic	Readings
Week 1	Introduction	Ch. 1
Week 2-3	Words, Text-to-speech	Excerpts from Ch. 3-4, Ch. 6-7
Week 4	Syntax: Part-of-speech tagging	Ch. 8
Week 5-8	Syntax: Parsing	Ch. 9-10, excerpts from Ch. 11-13
Week 9-10 (3/20-24)	Semantics	Ch. 14-16
	<b>Spring Break, no class</b>	
Week 11-12	Discourse and Dialogue Processing	Ch. 18-19
Week 13	Natural Language Generation	Ch. 20
Week 14-15	Applications: Machine Translation, NLP and the web	

## Important Dates

Note: homework and project deadlines are **tentative**.

Date	Event
2/2 (Th)	Homework 1 due
2/9 (Th)	<b>Midterm 1</b>
2/23 (Th)	Homework 2 due
3/16 (Th)	<b>Midterm 2</b>
3/30 (Th)	Project Part 1 due
4/27 (Th)	Project Part 2 due
Finals week (5/1-5)	<b>Final</b>

# Grading Criteria

The class will be graded out of 1000 points, distributed as follows:

- **4 quizzes** (40 points): Each quiz will take 10' and will be worth 10 points. Quizzes will be announced the lecture preceding the day in which they're given.
- **2 Homework Assignments** (100 points): Each homework will be worth 50 points.
- **Project** (300 points): The project will comprise two (possibly three) parts, for a total of 30% of the grade.
- **3 Exams** (560 points): 2 midterms (worth 140 and 160 points, respectively), 1 final (260 points).

**Important Note:** To pass the class you must get at least 60% of the total available points for the three exams, i.e, your cumulative score across the three exams must be at least 330 points.

Letter grades will be decided **only at the end**. However, the following guidelines will be adhered to:

Overall Score (undergraduate)	Overall Score (graduate)	Letter grade
88%	92%	A
78%	82%	B
68%	72%	C

## Policies on homeworks and exams

### General Policies

1. Late homeworks will not be accepted in any case, unless there is a **documented** personal emergency. Arrangements must be made with the instructor as soon as possible after the emergency arises, preferably before the homework due date.  
**Advice:** If for whatever reason you don't manage to finish an assignment, hand in what you have. Partial credit will be given.
2. Statute of Limitations: **Two weeks!** No grading questions or complaints — **no matter how justified** — will be listened to two weeks after the item in question has been returned.

### Homeworks

There will be 2 homeworks and a project. Homeworks will be mostly pen and pencil. The project (most likely in two, possibly in three parts) will consist of implementing a simple NLP systems. You will be given a choice of two or three programming languages in which to do your implementation.

Homeworks will have to be handed in via the web page. More details will be available later.

### Exams

1. The two midterms will be given during class time; consequently, **no make-ups** will be given.
2. Exams will be closed-book.
3. The final will be cumulative.

# Policy on Academic Integrity

**Academic dishonesty will not be tolerated.** Please see the CS department policy below on the topic; this policy specifies penalties for violations.

What is academic dishonesty? To hand in any work which is not 100% the student's creation, unless you are explicitly allowed to do so. Thus:

1. **Exams.** All work on all exams must be individually performed.
2. **Homeworks:** no student may give any other student any portion of their solutions or code, through any means. Students are not allowed to help each other debug the code, or to show each other any portions of code or homework.

**Important Note:** every semester somebody is caught red-handed and as a consequence fails the class. Isn't it better to get a B or a C than an E?

## CS department policy on academic dishonesty

The CS Department will not tolerate cheating by its students. The MINIMUM penalty for any student found cheating will be to receive an E for the course and to have the event recorded in a department and/or College record. The maximum penalty will be expulsion from the University.

We intend to devote more effort than in the past to detecting and punishing cheating. Cheating includes all the following, though this is not a complete list:

- Copying or any other form of getting or giving assistance from another student during any test, quiz, exam, midterm, etc.
- Plagiarism—turning in writing that is copied from some other source.
- Obtaining solutions to homework by posting to the Internet for assistance, purchasing assistance, obtaining copies of solutions manuals for instructors, and obtaining copies of previous year's homework solutions.
- Computer programs: Any time you look at another student's code, it is cheating. (Exception: If you are EXPLICITLY told that you may do so by the instructor, for instance, in working on a large group project in the 400-level software engineering course.)

For computer programs, if for some reason we cannot determine who copied from whom, we may, at our discretion, give failing grades to both students.

It is the responsibility of all engineering and computer science professionals to safeguard their company's "trade secrets." An employee who allows trade secrets to be obtained by competitors will almost certainly be fired. So, YOU are responsible for making sure that your Unix directories have permissions set so that only you can read your files, for being sure to log out at the end of working in the computer lab, etc.