

Natural Language Processing – Spring 2016
COMP 7/8780-PSYC 7780/8780
Vasile Rus

Contact Information:

Office: 323 Dunn Hall	Department Office: 375 Dunn Hall
Phone: 901-678-5259	Department Phone: 901-678-5465
E-mail: vrus@memphis.edu	

Office Hours:

Monday	Tuesday	Wednesday	Thursday	Friday
	10:00-11:00AM		10:00-11:00AM	
<i>Also by Appointment</i>				

Course Description:

COMP 7780-8780. (from the University catalog).

(Same as PSYC 7221-8221). Computational aspects, algorithms, and techniques for human language processing including lexical analysis, syntactic parsing, semantics, word sense disambiguation, logic forms, dialogue, and pragmatics; applications include question answering and information extraction among others. PREREQUISITE: COMP 6040 or 6041 or permission of instructor.

Why this course?

Since the advent of the Internet we are faced with an Information Overload problem. Most of the electronic information is in textual form or text combined with other media (tables, sounds, images, videos, etc.). There is a growing need for language processing tools. The Natural Language Processing class offers an overview of the techniques and tools available for language processing.

The class introduces computational techniques for understanding, learning and generating Natural Language with an emphasis on algorithms and software tools. It will cover tokenization - the separation of words from other items in written text such as commas, lexical semantics - the study of individual words, syntax - how words group together in more complex language structures, semantics - what the meaning of a sentence is, and pragmatics - what the meaning of a sentence is given the surrounding discourse context. The course will introduce both relational and statistical approaches to NLP, expose students to state-of-the-art tools and technologies and their impact on a variety of applications. It will also provide insights to open problems in the field.

Resources:

See the class website: <http://www.cs.memphis.edu/~vrus/teaching/nlp/>

Required Text

D. Jurafsky and J. Martin: *Speech and Language Processing, 2nd Edition, Prentice Hall*

Recommended Texts

C. Manning and H. Schütze: *Foundations of Statistical Natural Language Processing*

Other Resources:

See the class website: <http://www.cs.memphis.edu/~vrus/teaching/nlp/>

Evaluation:

The University policy requires to email a grade to a student's U of M email address *only*.

Final Grades:

Homework 30%, Project 35%, Midterm 15%, Final 15%, Class Participation 5%

PhD Students are required to make an extra paper presentation as part of the Homework grade. The presentation consists of power point slides presented in front of the class for 15 minutes and an accompanied report of 3-5 pages (single spaced).

Grading Scale:

<i>Grade</i>	<i>Letter Grade</i>
90-100+	A
80-89	B
70-79	C
60-69	D
0-59	F

2.5 above or below the cut-off will earn you a + or – in front of your grade.

For example: 89 has a letter equivalent of B+

Exception: A- is for 90-91, A for scores ranging from 92 to 97, anything above 98 leads to A+

Course Policies:

Attendance

Students are strongly encouraged to attend all lectures. **Active participation** to class discussions counts toward your final grade.

Late Policy

Students will have on average one-two weeks from the date the work is assigned. Late submissions are not accepted. In exceptional cases you may have a 48-hour grace period at the cost of 50% of the grade (Students **must** ask for it before the due date).

Testing Policy

Usually exams are closed books. There are no make-up exams. Any code developed as part of the class work must follow the **coding-style** guidelines described on the web site. The coding-style will be strictly enforced.

Plagiarism/Cheating Policy:

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and **will not be tolerated**. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature (including the internet) on their assignments, but **appropriate references must be included for the materials consulted**, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the University Judicial Affairs Office for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to:

<http://www.people.memphis.edu/~jaffairs/>

Course Syllabus (tentative)

Week 1: Introduction, Words, Morphology, Stemming

Week 2: Stemming, Part-of-Speech Tagging

Week 3: Introduction to Perl

Week 4: Corpus-based Methods, Language Modeling, N-grams

Week 5: HMM tagging

Week 6: Syntax and Grammar, CFG and Parsing Algorithms

Week 7: Probabilistic CFG and parsing

Week 8: NO-CLASS – Spring Break!

Week 9: MIDTERM, Unification Grammar, Lexical Semantics

Week 10: Lexical Semantics, LSA

Week 11: Logic Form

Week 12: Discourse Processing

Week 13: Applications: Information Extraction, Question Answering, Summarization, Information Retrieval and Web Search

Week 14: Applications: Information Extraction, Question Answering, Summarization, Information Retrieval and Web Search

Week 15: Review, Discussions

Week 16: Final Exam