CS4650/CS7650 – Natural Language
Spring 2010 Syllabus

Instructor: David Novick
Office Hours: TBD

Overview This course introduces the theory and methods of natural language processing (NLP). NLP systems understand and produce human language for applications such as information extraction, machine translation, automatic summarization, question-answering, and interactive dialog systems. The course covers knowledge-based and statistical approaches to language processing for syntax (language structures), semantics (language meaning), and pragmatics/discourse (the interpretation of language in context). For graduate students, the course will also cover aspects of current research in NLP.


Course Outcomes
Upon successful completion of this course, students will be able to demonstrate accomplishments of knowledge and comprehension, application and analysis, and synthesis and evaluation:

1. Knowledge and Comprehension

Explain the elements and applications of:

- Part-of-speech tagging
- Parsing
- Semantic analysis
- Machine learning of natural language
- Discourse processing
- Dialog models
- Natural language applications
2. Application and Analysis

Apply skills of:

- Programming in Python
- Simple language processing with the Natural Language Tool Kit (NLTK)
- Simple machine learning with the NLTK
- Apply cooperative group skills and critical-thinking skills germane to natural language processing
- Understand and explain research papers in natural language processing (graduate students)

3. Synthesis and Evaluation

- Compare and contrast approaches to natural language processing
- Discuss the limitations and promise of NLP

Standards of Conduct

You are expected to conduct yourself in a professional and courteous manner, as prescribed by the Georgia Tech Student Code of Conduct. Graded work, such as homework and tests, is to be completed independently and should be unmistakably your own work, although you may discuss your project with other students in a general way. You may not represent as your own work material that is transcribed or copied from another person, book, or any other source, e.g., a Web page. The instructor is required to—and will—report academic dishonesty and any other violation of the Standards of Conduct to the Dean of Students.

Disabilities

If you feel that you may have a disability that requires accommodation, contact the Access Disabled Program for Tech Students (ADAPTS), http://www.adapts.gatech.edu, adapts@vpss.gatech.edu.

Assignments

Reading and homework assignments will be handed out or announced in class. If you miss a class, it is your responsibility to find out what you missed. You should expect to spend at least seven hours per week outside of class on reading and homework.
Grading
Your semester grade will be based on a combination of programming assignments in Python and the NLTK, two midterm exams, and a final exam. Graduate students will read and present papers from the research literature, and will write and present a research paper. The approximate percentages are as follows:

Undergraduate:

- 40% Programming assignments
- 30% Two midterms
- 30% Final exam

Graduate:

- 30% Programming assignments
- 20% Two midterms
- 25% Research paper
- 5% Research literature presentation
- 20% Final exam