

University of North Texas  
CSCE5290-001 / CSCE 4930-004, Fall 2016  
Natural Language Processing

**Meetings** Mondays, Wednesdays and Fridays, 01:30pm–02:20pm, NTDP B142  
**Instructor** Dr. Eduardo Blanco  
**Office** NTDP F245  
**Email** [eduardo.blanco@unt.edu](mailto:eduardo.blanco@unt.edu)  
**TA** Nishitha Guntakandla, [NishithaGuntakandla@my.unt.edu](mailto:NishithaGuntakandla@my.unt.edu)

### Textbook

Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (second edition). Prentice Hall. Website: <http://www.cs.colorado.edu/~martin/slp.html>

### Course Contents

This course provides an introduction to natural language processing (NLP) covering morphology, syntax, semantics and pragmatics. The fundamental algorithms and techniques in each of these areas are studied, and existing resources and tools presented. The course introduces both knowledge-based and statistical approaches to NLP, illustrates the use of NLP techniques and tools in several applications, and provides insight into many open research problems.

Topics include:

- Regular expressions and automata
- N-grams, language models
- Part-of-speech tagging
- Spelling correction
- Hidden Markov Models, forward and Viterbi algorithms
- Formal grammars, syntactic parsing
- Lexical semantics, word sense disambiguation, semantic role labeling
- Knowledge representation from text
- Information extraction
- Special topics: open problems and current research

You are expected to check <https://learn.unt.edu/> often for course material, homework assignments and grades.

### Prerequisites

CSCE 3110: Data Structures and Algorithms.

### Grading

Homework	40%
Quizzes	10%
Project	15%
Midterm	15%
Final exam	20%

- There will be five homework assignments. Homework grade will be the average of the four highest-graded homework assignments. Homework is due at the beginning of class at least a

week after it is assigned. Homework assignments will include both written and programming exercises. As a rule, late homework will not be accepted.

- There will be five quizzes. Quiz grade will be the average of the four highest-graded quizzes.
- A list of project ideas will be distributed to the class.
- The midterm exam will be during class on TBD.
- The final exam will be on Saturday, December 10, 2016, 1:30pm–3:30pm.

### **ABET outcomes**

- Define and evaluate regular expressions.
- Understand n-grams, language models and smoothing techniques.
- Understand and solve the problem of part-of-speech tagging.
- Understand Hidden Markov Models and the Viterbi algorithm.
- Define formal grammars and recognize the language accepted by a grammar.
- Understand basic syntactic parsing algorithms (CKY, Early, etc.) and statistical parsing.
- Understand some problems in computational semantics, for example, semantic role labeling, lexical semantics, word sense disambiguation.
- Familiarity with knowledge-based and statistical approaches in NLP.

### **Academic Integrity**

Academic Integrity is defined in the UNT Policy on Student Standards for Academic Integrity. Any suspected case of Academic Dishonesty will be handled in accordance with the University Policy and procedures. Possible academic penalties range from a verbal or written admonition to a grade of F in the course. Further sanctions may apply to incidents involving major violations. You will find the policy and procedures at: <http://vpaa.unt.edu/academic-integrity.htm>.

Each topic discussed in class will have associated homework. Students may discuss homework problems and approaches with each other, but must write their solutions individually. Students may not copy homework from any source, including other students or the internet. No collaboration is allowed in quizzes and exams.

### **Religious Observance**

In accordance with state law, a Student absent due to the observance of a religious holiday may take examinations or complete assignments scheduled for the days missed, including those missed for travel, within a reasonable time after the absence. Students should notify the instructor in each course of the date of the anticipated absence as early in the semester as possible. Only holidays or holy days observed by a religion whose place of worship is exempt from property taxation under Section 11.20 of the Tax Code may be included. A student who is excused under this provision may not be penalized for the absence, but the instructor may appropriately respond if the student fails satisfactorily to complete the assignment or examination.

Check <http://policy.unt.edu/policy/15-2-5> for more information.

### **Disability Accommodations**

The University of North Texas makes reasonable academic reasonable accommodation for students with disabilities. Students seeking reasonable accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request reasonable accommodations at any time, however, ODA notices of reasonable accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of reasonable accommodation for every semester and must meet with each faculty member prior

to implementation in each class. Students are strongly encouraged to deliver letters of reasonable accommodation during faculty office hours or by appointment. Faculty members have the authority to ask students to discuss such letters during their designated office hours to protect the privacy of the student. For additional information see the Office of Disability Accommodation website at <http://www.unt.edu/oda>. You may also contact them by phone at 940.565.4323.