Course Syllabus: Intro to Natural Language Processing

Course Information
CS 4301 – Section 001
Time: M/W 2:30 pm – 3:45 pm
Location: SLC 2.302

We will use Piazza for class discussion, announcements, exam reminders, and homework instructions and due dates. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza so that everyone has the same information.
Sign up here: https://piazza.com/utdallas/fall2017/cs4301

Piazza also has a free android/iphone app.

Professor Contact Information
Dr. Karen Mazidi
Karen.Mazidi@utdallas.edu
Office: ECSS 3.203
Office phone: 972-883-3868
Office hours: M/W 4:00 – 5:00 pm

TA Information
TBD

Course Description
(3 semester credit hours) This course covers state-of-the-art methods for natural language processing. After an introduction to the basics of syntax, semantic, and discourse analysis, the focus shifts to the integration of these modules into natural-language processing systems. In addition to natural language understanding, the course presents advanced material on lexical knowledge acquisition, natural language generation, machine translation, and parallel processing of natural language.

Student Learning Objectives
After successful completion of this course, students will:

1. The ability to perform morphological analysis
2. The ability to compute n-gram statistics and to perform smoothing
3. Knowledge about parts-of-speech and ability to use and understand part-of-speech taggers
4. Knowledge about various parsing algorithms and parsing techniques
5. Knowledge of semantic representations including FrameNet, WordNet, PropBank and semantic parsing techniques that employ them
6. Knowledge about statistical machine translation
7. Applying the NLP techniques to modern applications

UTD

Required Textbooks and Materials

Speech and Language Processing by Daniel Jurafsky and James H. Martin
The 2nd edition is on reserve in the library
The 3rd edition is still being written, some chapters are available:
https://web.stanford.edu/~jurafsky/slp3/

NLTK Book: http://www.nltk.org/book/

The following resources may be helpful if you are new to Python:

- Think Python online version (beginner resource)
- http://www.diveintopython3.net/ (a more advanced introduction)

Topics and Sequence

- Week 1 – 8/21 – Read: J&M2:1, NLTK:1,2
  - course intro
  - Python
- Week 2 – 8/28 – Read: J&M3:2
  - Python and NLTK
  - **Homework 1: File I/O and FreqDist**
- Week 3 – 9/4 – Read: J&M3:4
  - Monday 9/4 Labor Day no classes
  - Linguistics 101
  - **Homework 2: Morphology**
- Week 4 – 9/11 – Read: J&M3:10
  - collocations and corpora
  - POS tagging
  - **Homework 3: N-gram Language Model**
- Week 5 – 9/18 – Read: J&M3:10
  - Exam 1
  - review Exam 1
- Week 6 – 9/25 – Read J&M3:29,30
  - chatbots
  - CFG grammar
  - syntax parsing and trees
  - **Homework 4: Drawing Syntax Trees**
- Week 7 – 10/2 – Read J&M3:22,15
  - semantic role labeling
  - vector semantics
- Week 8 – 10/9 – Read: J&M3:17
  - word sense disambiguation
  - semantics

- **Homework 5: ACL paper review**
- Week 9 – 10/16 – Read J&M3:18
  - sentiment detection
  - **Homework 6: SentiWordNet**
- Week 10 – 10/23 – Read:
  - Exam 2
  - Project discussion/preliminary results
- Week 11 – 10/30 – Read: J&M3:6
  - intro to ML and sklearn, numpy, pandas
  - naïve bayes classifiers
- Week 12 – 11/6 – Read: J&M3:7
  - logistic regression
  - SVM
  - **Homework 7: Language ID with NB, SVM**
- Week 13 – 11/13 – Read: J&M3:8
  - neural networks
  - deep learning
  - **Homework 8: optional ML adventure**
- Week 14 – 11/20
  - Fall break – no classes
- Week 15 – 11/27 – Read: Chapter 6
  - Exam 3
  - review exam
  - optional Homework 8 presentations
- Week 16 – 12/4
  - Project Presentations
- Final Exam Week: Fri 12/8 – Thurs 12/14

Grading

Course grade is composed of:

- 25% - homework; approximately 8 assignments
- 45% - 3 exams; 15% each
- 25% - project
- 5% - attendance (1 - %unexecused absences)

Letter grades will be assigned according to the UTD +/- conventions.

Course & Instructor Policies

- Students should take notes by hand using either physical or electronic paper. Here is why: http://www.scientificamerican.com/article/a-learning-secret-don-t-take-notes-with-a-laptop/
- During class coding time when laptops are allowed, please be mature, be respectful of the distractions that you may cause others if you are off task.

Perhaps you think you can multi-task, but science says otherwise: http://www.npr.org/2013/05/10/182861382/the-myth-of-multitasking

- Assignments must be turned in on the due date, by midnight. It's a good idea to screen-shot your upload in eLearning. This is the only way you can prove that you really did turn in the homework.
- Late assignments are deducted by 10% on the first and second days late. After two days, the assignment will not be accepted.
- Makeup exams are not given unless prior permission has been granted due to extenuating circumstances.
- Do not turn in programming code or other work that is not your own. This will result in a zero for all parties. Discussion and collaboration are good things, turning in someone else's work as your own is not.
- **If you do not agree with a grade you have been given, you must make your case within a week of receiving the grade or the grade stands as is.**

Comet Creed

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

"As a Comet, I pledge honesty, integrity, and service in all that I do."

Attendance Policy

Per Computer Science administration guidelines, please be aware that ***3 consecutive absences leads to one letter grade drop. Four consecutive absences lead to an***

UT Dallas Policies and Procedures

The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus.

Please go to http://go.utdallas.edu/syllabus-policies for these policies.

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.