# NYUAD Course CS-AD 220 – Spring 2016

# Natural Language Processing

- **Instructor**: Prof. Nizar Habash
- **Office**: CR-189 (A2-189)
- **Email**: nizar.habash@nyu.edu
- **Website**: www.nizarhabash.com

- **Class Hours**: Tue/Thu 2:10 - 3:25
- **Class Location**: C2-E049
- **Office Hours**: By appointment

## Slides and Handouts

All slides and handouts will be available online HERE.

## Description

The field of natural language processing (NLP), also known as computational linguistics, is interested in the modeling and processing of human (i.e., natural) languages. This course covers foundational NLP concepts and ideas, such as finite state methods, n-gram modeling, hidden Markov models, part-of-speech tagging, context free grammars, syntactic parsing and semantic representations. The course will survey a range of NLP applications such as information retrieval, summarization and machine translation. Concepts taught in class will be reinforced in practice by hands-on assignments.

## Text Books

**J+M:** Daniel Jurafsky and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." (2nd Edition). Pearson Prentice Hall, 2008. **The book is available at the bookstore.**

**NH:** Nizar Habash. "Introduction to Arabic Natural Language Processing." Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2010. This book is available online (click HERE).

## Requirements

**Components of the Final Grade:** There will be four homework assignments and two exams (one mid-term and one final). The assignments are varied in form including on-paper exercises (such as walking through algorithms), hands-on programming, building systems using off-the-shelf tools, and data analysis. Assignments count 55% (one earns 10% and three earn 15% each); exams count 40% (20% mid-term and 20% final); and class participation (asking questions and participating in discussions that demonstrate having read the assigned readings) earns 5%.

**Participation and Attendance Policies:** Attendance required. Every unexcused absence (i.e., not pre-cleared with instructor, or not with medical documentation provided to instructor) will automatically cause one grade level drop. Participation is expected in class discussions of specific NLP problems and solutions in a way that demonstrates preparedness.

**Assignment Submission Policies:** All assignments are due at 11:59pm on the due date specified in syllabus. For late submissions, 10% will be deducted from the homework grade per each late day.

**Extra Credits:** There are multiple opportunities to earn extra credit in the class (up to 20% extra!):

1. 1% (of the whole grade) for *each* NLP CS Colloquium talk attended (11am Thursdays) -- up to 5% total. The student must submit a one page summary of the talk by the week after the talk. The expected dates for NLP talks in Spring 2016 are
    - 18-Feb Prof. Khaled Shaalan
    - 17-Mar Prof. Dough Oard
    - 31-Mar Prof. Mona Diab
    - 14-Apr Prof. Hend Alkhalifa
    - 21-Apr Prof. Jan Hajic
2. Up to 5% (of the whole grade) for participation in the NYUAD Hackathon on Social Good in the Arab World (APril 15-17, 2016). The student earns 2% for participating in a team working on an NLP related component. If the team the student is in wins, extra points are given (+1% third place; +2% second place; +3% first place).
3. Up to 5% (of the whole grade) for pop-up quizzes.
4. Up to 5% (of the whole grade) will be awarded as part of the last assignment on Machine Translation (stay tuned).

## Academic Integrity

As set forth in NYU Abu Dhabi's Academic Integrity Policy, the relationship between students and faculty at NYU Abu Dhabi is defined by a shared commitment to academic

excellence and is grounded in an expectation of fairness, honesty, and respect, which are essential to maintaining the integrity of the community. Every student who enrolls and everyone who accepts an appointment as a member of the faculty or staff at NYU Abu Dhabi agrees to abide by the expectation of academic honesty. The full policies and procedures relating to Academic Integrity may be found on the NYUAD Student Portal.

## Class Schedule

| Date | Topic | Reading | Other |
|---|---|---|---|
| Thu 28th Jan | Introduction to NLP | None | |
| Tue 2nd Feb | Introduction to NLP | J+M Chap 1 | |
| Thu 4th Feb | Regular Expressions and Basic Text Processing | J+M Chap 2 (intro,2.1) | Assignment 1: due by Session 7 (Feb 18 midnight) |
| Tue 9th Feb | Regular Expressions and Basic Text Processing | Handout (Regex Cheat Sheet) | |
| Thu 11th Feb | Finite State Automata | J+M Chap 2 (2.2) | |
| Tue 16th Feb | Finite State Automata | J+M Chap 2 (2.3 to end) | |
| Thu 18th Feb | Morphology and Finite State Transducers | J+M Chap 3 (intro up to 3.8); | Assignment 2: due by Session 13 (Mar 10 midnight) |
| Tue 23rd Feb | Morphology and Finite State Transducers | J+M Chap 3 (3.8 to end); NH Chap 4 | |
| Thu 25th Feb | Language Modeling | J+M Chap 4 (intro up to 4.5) | |
| Tue 1st Mar | Language Modeling | J+M Chap 4 (4.5 up to 4.9) | |
| Thu 3rd Mar | Part-of-Speech Tagging | J+M Chap 5 (intro up to 5.5) | |
| Tue 8th Mar | Part-of-Speech Tagging | J+M Chap 5 (5.5 up to 5.8) | |
| Thu 10th Mar | Part-of-Speech Tagging | J+M Chap 5 (5.8 to end); handout (Pasha et al., 2014) | |
| Tue 15th Mar | Midterm | All previous readings | |
| | SPRING BREAK | | |
| Sat 26th Mar | Syntax and Parsing | J+M Chap 12 | Assignment 3: due by Session 21 (Apr 14 midnight) |
| Tue 29th Mar | Syntax and Parsing | None | |
| Thu 31st | | | |

| | | | |
|---|---|---|---|
| Mar | Syntax and Parsing | J+M Chap 13 | |
| Tue 5th Apr | Syntax and Parsing | None | |
| Thu 7th Apr | Machine Translation | J+M Chap 25 (intro up to 25.5) | |
| Tue 12th Apr | Machine Translation | Handout (Papineni et al., 2002) | |
| Thu 14th Apr | Machine Translation | J+M Chap 25 (25.5 to end); | Assignment 4: due by Session 27 (May 10) |
| Tue 19th Apr | Machine Translation | Handout (Zens et al., 2002) | |
| Thu 21st Apr | Machine Translation | NH Chap 8 | |
| Tue 26th Apr | Machine Translation | Handout (Habash+Sadat, 2006) | |
| Thu 28th Apr | Lexical Semantics | J+M Chap 19 | |
| Tue 3rd May | Lexical Semantics | J+M Chap 20 | |
| Thu 5th May | No Class - Isra and Miraj Holiday | | |
| Tue 10th May | Information Extraction | J+M Chap 22 | |
| Thu 12th May | Question Answering and Summarization | J+M Chap 23 | |
| | FINAL EXAM TBD | | |