# ECE 684/IDS 703: Natural Language Processing

# Fall 2022

*Instructor:*
Patrick Wang
patrick.wang@duke.edu
Gross 266

*Lecture:*
Wilkinson 021
Mon/Wed 5:15pm–6:30pm ET
Zoom

## Overview

This course discusses applications of natural language processing and explores algorithmic and analytic methods specific to natural language data. Applications include document classification, speech recognition, and machine translation. This course covers both statistical methods (HMM, LDA, ...) and "black-box" methods (RNN, word2vec, ...).

## Learning Objectives

At the conclusion of this course, students will be able to:

- identify applications of natural language processing,
- discuss the potentials approaches and challenges for specific applications,
- develop general algorithmic workflows for specific applications, and
- implement NLP systems using a high-level programming language.

## Grading

- homework assignments (60%)
- final project (40%)

**Homework assignments** will be posted on Sakai and should be turned in on Sakai. Late submissions will immediately incur a 50% penalty unless approved by the instructor ahead-of-time.

**Final project** is an independent NLP project designed by students. Students may work in groups of up to 3. Written reports will be due on Dec 19 at 12:00pm EST, in lieu of the final exam.

## Tentative Schedule

| **date** | topic | readings |
|---|---|---|
| **8/29** | intro & Python | |
| **8/31** | regular expressions | Jurafsky and Martin 2.1 |
| **9/5** | tokenization | Jurafsky and Martin 2.4 |
| **9/7** | edit distance & noisy-channel model | Jurafsky and Martin 2.5, B |
| **9/12** | stemming | Jurafsky and Martin 2.4.4 |
| **9/14** | language modeling & n-gram model | Jurafsky and Martin 3 |
| **9/19** | sentiment analysis & Naïve Bayes | Jurafsky and Martin 4, 21 |
| **9/21** | part-of-speech tagging & HMM | Jurafsky and Martin 8, Rabiner (1989) |
| **9/26** | named entity recognition | |
| **9/28** | topic modeling & LDA | Blei et al. (2003) |
| **10/3** | TF-IDF & PMI | Jurafsky and Martin 6 |
| **10/5** | context vectors (SVD) | Jurafsky and Martin 6 |
| **10/12** | feed-forward neural networks | Jurafsky and Martin 7 |
| **10/17** | autoencoders (word2vec) | Jurafsky and Martin 6.8 |
| **10/19** | recurrent neural networks | Jurafsky and Martin 9 |
| **10/24** | conditional random fields | Sutton (2012) 2 |
| **10/26** | gated networks | Jurafsky and Martin 9.4 |
| **10/31** | encoder-decoder networks | Jurafsky and Martin 10, Sutskever et al. (2014) |
| **11/2** | attention mechanisms | Bahdanau et al. (2016) |
| **11/7** | transformer architectures | Vaswani et al. (2017) |
| **11/9** | training & evaluation | |
| **11/14** | training & evaluation (2) | Peters et al. (2018), Devlin et al. (2019) |
| **11/16** | big models & compression | |
| **11/21** | big models & finetuning | |
| **11/28** | structural parsing | Jurafsky and Martin 12, 13, 15 |
| **11/30** | question-answering systems | Jurafsky and Martin 25 |
| **12/19** | *final project due* | |

## Policies

Students in this course are expected to abide by the commitments they made in signing the Duke Compact to protect the health and safety of their fellow students, faculty, staff, families and neighbors. First time, minor violations of COVID-19 conduct expectations will be met with appropriate educational responses. However, anyone who fails to comply with the expectations of the Duke Compact more than once, or who flagrantly commits a serious violation that creates a health or safety risk to others in the Duke community, will be subject to more significant consequences, beginning with loss of the privilege to attend courses in-person and/or loss of access to campus, and moving up to suspension or expulsion.

# Readings


Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv:1409.0473 [cs, stat]*, May 2016. URL `http://arxiv.org/abs/1409.0473`. arXiv: 1409.0473.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019. URL `http://arxiv.org/abs/1810.04805`. arXiv: 1810.04805.

Dan Jurafsky and James H Martin. Speech and language processing, 2019. URL `https://web.stanford.edu/~jurafsky/slp3`.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv:1802.05365 [cs]*, March 2018. URL `http://arxiv.org/abs/1802.05365`. arXiv: 1802.05365.

Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. Publisher: Ieee.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to Sequence Learning with Neural Networks. *arXiv:1409.3215 [cs]*, December 2014. URL `http://arxiv.org/abs/1409.3215`. arXiv: 1409.3215.

Charles Sutton. An Introduction to Conditional Random Fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373, 2012. ISSN 1935-8237, 1935-8245. doi: 10.1561/2200000013. URL `http://www.nowpublishers.com/article/Details/MAL-013`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances*

*in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.