# 2110594: Natural Language Processing

## Mondays 9:00-12:00

ผู้สอน อ.พีรพล และ อ.เอกพล


Course Outline

The course will give an overview of the Natural Language Processing techniques in the transition period between traditional techniques and deep learning. The course will touch upon various standard NLP tasks, such as tokenization, language modeling, semantics, part-of-speech tagging, and parsing. Each topic will discuss both conventional and deep learning techniques. The second part of the course will go into application domains such as document classification, question answering, and chatbots. Many of the assignments will be in the context of Thai language. This course is not meant to be an entry-level machine learning course, and will not cover the basics of machine learning. This course is recommended for 4th year students who have already taken some machine learning course, and 3rd year students who are interested in doing senior projects that are related to NLP.

Tools: Python, keras, ntlk, scikit-learn

**Prerequisites: some background in machine learning/data science**

Schedule (กำหนดการอาจมีการเปลี่ยนแปลง)

| คาบเรียนที่ | เนื้อหา | การบ้านและควิซ |
|---|---|---|
| 1 - 8/1 | Intro to NLP (P) | HW0-Hello GCloud |
| 2 - 15/1 | Tokenization<br>LexTo, DNN, CNN, LSTM with Keras. (E) | HW1-TUMKUDv2 |
| 3 - 29/1 | Language modeling (P)<br>N-grams, smoothing, Neural LM | HW2-LM |
| 4 - 5/2 | Representation (P)<br>TF-IDF, word and sentence embeddings, adaptation | HW3-Embeddings |
| 5 - 12/2 | PoS tagging and information extraction (E)<br>CRF and beam search | HW4-CRF |
| 6 - 19/2 | Parsing (E)<br>PCFG, Recursive neural networks | Study for midterm |
| 7 - 26/2 | In class midterm | |
| | Midterm exam week | |
| 8 - 12/3 | Document/sentiment classification (E)<br>LDA, Naive Bayes, EM, LDA2Vec | HW5-Intent classification |
| 9 - 19/3 | Question Answering, Reading comprehension, Text generation (P)<br>Sequence2Sequenc models, Attention models | HW6-Text Generation |
| 10 - 26/3 | Guest speakers session 1<br>Sertis: NLP in industry, elastic search, etc.<br>Aj. Wirot: Corpus creation considerations | HW7-Project scoping |
| 11 - 2/4 | ASR in NLP and spoken dialogue systems (E)<br>Guest speakers session 2<br>BoT: Conversational agent<br>HBot: HBot API workshop | Project start |
| 12 - 9/4 | Spell correction and recent advances in NLP (E)<br>Noisy channel models, transformer, GAN | |
| 13 - 16/4 | Pytorch tutorial<br>Paper presentation | |
| 14 - 23/4 | Project presentation | |

การส่งงานสาย
การส่งงานสายเกิน5นาทีหลังจากหมดกำหนดส่งจะไม่ได้คะแนนในทุกกรณี

เกณฑ์การวัดผล
Assignments 30% (5% each with 5% extra)
Midterm 35%
Project 35%

การตัดเกรด
> 80% A
> 75% B+
> 70% B
> 65% C+
> 60% C
> 55% D+
> 50% D
< 50% F

หนังสือเรียน
ไม่มีหนังสือเรียนบังคับ