



ITMO UNIVERSITY

NLP Basic and Selected Topics

A Practical and Easy Introduction to Selected Topics

Aug 31st, 2019

Overview of the Unit Today

- 1) Applications of NLP / Introduction (30min)
- 2) Practical NLP (NLTK / pythainlp) (45min)**
- 3) Modern NLP with ML/DL (45min)
- 4) Example: Word Similarity and WordNet (30min)
- 5) Modern NLP with fastAI / flair (30min)

Overview

- ✓ Introduction to NLTK
- ✓ Practical usage in Google Colab
- ✓ Introduction to pythainlp
- ✓ Practical Usage in Google Colab

NLTK

- ✓ Python library for traditional NLP
- ✓ Exists for a long time
- ✓ Good for getting started and doing easy things

Overview

- ✓ Look at other slideset (about NLTK)
- ✓ Look at code on CoLab

Exercise

Open Google Colab (you need to be logging into Google)

colab.research.google.com

- 1.) Create a string (a sentence).
- 2.) Count how many words are in the sentence.
- 3.) Apply part-of-speech tagging
- 4.) Plot the frequency distribution
- 5.) Optional: do sentiment tagging of the sentence

pythainlp

- ✓ Look at other slideset (about NLTK)
- ✓ Look at code on CoLab
- ✓ <https://github.com/PyThaiNLP/pythainlp/blob/dev/notebooks/pythainlp-get-started.ipynb>

Exercise

Use pythainlp

- 1.) Create a Thai string (a sentence).
- 2.) Do word splitting
- 3.) Romanize
- 4.) Apply part-of-speech tagging
- 5.) Apply Named-Entity-recognition

deepcut

- ✓ Does wordsplitting
- ✓ Should have better accuracy
- ✓ See colab