

Technologies and Implementation of Natural Language / 自然語言技術與實作

CJ Wu/吳齊人

Summer 2021 Syllabus

This introductory course is about a variety of ways to represent human languages for Computer Science (CS) undergraduate students. This course will introduce students to the basics of Natural Language Processing (NLP). The students will gain the ability to exploit languages representations to write programs based on the modern data-driven techniques. This course will include some ideas, including Machine Learning and the recent Deep Learning. There will be an emphasis on rapid prototyping, a useful skill in many other areas.

Topics

Topics covered include:

- Web crawling and indexes
- Language modeling
- Representation learning
- Word Embeddings
- Text classification
- Sequence modeling
- Machine learning models
- Deep learning models

Textbook

here is no required textbook.

References

1. Jurafsky and Martin,. Speech and Language Processing. 3 edition
https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf
2. Raghavan, and Schutze. 2008. Introduction to Information Retrieval. Cambridge University Press.
<https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
3. Natural Language Processing With Python's NLTK Package.
<https://realpython.com/nltk-nlp-python/>
4. Tencent AI LAB (中文)
<https://ai.tencent.com/ailab/zh/paper/?page=1>
5. CKIP Lab (中文)
<https://ckip.iis.sinica.edu.tw/>
6. Kaggle data
<https://www.kaggle.com/>

Grading

Your course grade will be determined approximately as follows:

- 40%: Final project
- 25%: Quizzes
- 35%: Exercises