

# Course Information

## Short Intro

Natural Language Processing (NLP) is a crucial part of Artificial Intelligence (AI), which modeling how people communicates to each other. The objective of this course is to provide a complete introduction to natural language processing techniques and their applications, especially in the era of deep machine learning approaches.

## Description

Natural Language Processing (NLP) is a crucial part of Artificial Intelligence (AI), which applies both Computer Science and Linguistics methodologies. NLP is sometimes referred to as Computational Linguistics (CL) when the speaker emphasizes more on linguistic structures. NLP is widely considered as the fundamental instrument of the information age, since applications facilitate people communicating in various kinds of language: web search, advertising, language translation, etc. The objective of this course is to provide a complete introduction to natural language processing techniques and their applications, as a first step leading towards more specialized graduate-level topics. In recent years, Deep Learning approaches have greatly improved the performance of almost every AI task, so modern methodologies using Deep Learning for NLP will be provided as the main theme.

The course will touch on the following topics:

- What is NLP? What can be counted as NLP?
- How is a sentence constructed grammatically?
- Why Word Segmentation is important for Chinese language?
- How to model language in traditional Statistics?
- How to represent words in a computer-friendly way?
- How does the most basic Neural Network work for NLP tasks?
- What are the benefits of using modern Deep Learning approaches?
- How to apply Deep Learning approaches in Machine Translation tasks?

Concepts will be illustrated with examples in the PyTorch framework.

**Keywords:** Natural Language Processing (NLP), Deep Neural Networks, PyTorch.

## Prerequisites

**Required:**

- College Calculus, Linear Algebra.
- Basic Probability and Statistics.
- Sufficient understanding of Programming.

**Recommended:**

- Numerical Optimization.
- Proficiency in Python.
- Knowledge of Machine Learning, Deep Learning.

## Teaching plan

1. Introduction 导言, Basic Text Processing 文本处理基础
2. Syntactic Analysis 句法分析, N-gram Language Models N元语法模型
3. Naïve Bayes 朴素贝叶斯, Logistic Regression 逻辑回归
4. Sequence Labeling 序列标注, Parts of Speech 词类, Named Entities 命名实体
5. Vector Semantics 向量语义, Word Embeddings 词嵌入

Computer Science - UG3/G	
Course	Natural Language Processing I
Term	2022F
Final	Tba
Credits	2
Staff	WU Xiaokun 吴晓坤
Lecture	32 hours (8 weeks)
计算机专业选修课 人工智能专业核心课	
课程名称	自然语言处理
授课时间	2022年春
考试形式	考试/考查
学分	2
讲者	吴晓坤
总计时长	32学时 (8周)

6. Recurrent Neural Networks 循环神经网络, Sentiment Analysis 情感分析
7. Modern RNNs 现代循环神经网络, Machine Translation 机器翻译

Several special topics:

1. Lexicon Parsing 词典分词, Chinese Word Segmentation 中文分词
2. Constituency Parsing 构成解析, Dependency Parsing 依存解析
3. Information Retrieval 信息检索
4. Encoder-Decoder 编码器-解码器, seq2seq 序列到序列学习
5. Attention 注意力机制, Transformer, BERT

## Schedule

Friday, Multiple locations.

Week	Date	Lecture	Handouts
1	2022/03/11	[ <a href="#">导言</a> ], [ <a href="#">文本处理</a> ]	[ <a href="#">NED</a> ]
2	2022/03/18	[ <a href="#">句法分析</a> ], [ <a href="#">词典分词</a> ]	[ <a href="#">评测</a> ]
3	2022/03/25	[ <a href="#">N元语法</a> ]	
4	2022/04/01	[ <a href="#">平滑处理</a> ], [ <a href="#">朴素贝叶斯</a> ]	
5	2022/04/08	[ <a href="#">朴素贝叶斯</a> ], [ <a href="#">序列标注</a> ]	[ <a href="#">逻辑回归</a> ]
6	2022/04/15	[ <a href="#">向量语义</a> ]	[ <a href="#">课程考核说明</a> ]
7	2022/04/22	[ <a href="#">词嵌入</a> ]	[ <a href="#">BPE</a> ]
8	2022/04/29	[ <a href="#">循环神经网络</a> ]	[ <a href="#">现代循环神经网络</a> ]

## Evaluation

- Attendance & participation: 20%.
- Understanding of the course: 20%.
- Final project: 50%.
- Honorable bonus: 10%.

课程考核说明: [[pdf](#)]

- release date: 2022/04/15
  - collect feedbacks: 1 week
- due: 2022/06/10 (2 weeks after course s).

## Textbook

Not mandatory but recommended:

- Jurafsky and Martin, *Speech and Language Processing*.
- Manning et al. *Introduction to Information Retrieval*.
- 何晗, 《自然语言处理入门》。

## Resources

- [Speech and Language Processing \(3rd ed. draft\)](#)<sup>1</sup>
- [CS224n: Natural Language Processing with Deep Learning](#)<sup>2</sup>

1. <https://web.stanford.edu/~jurafsky/slp3/><sup>1,2</sup>

2. <https://web.stanford.edu/class/archive/cs/cs224n/><sup>2</sup>