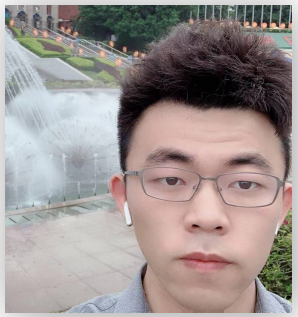
COMP5423 Natural Language Processing

Course Introduction

Course Introduction

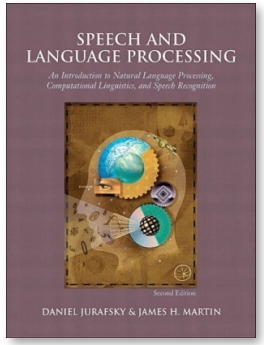
- **Lecturer (Responsible for Lecture Classes)**
  - Maggie, Wenjie Li
  - cswjli@comp.polyu.edu.hk
  - PQ707

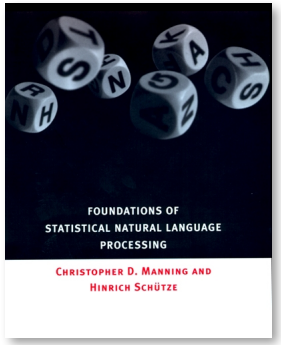
- **Tutors (Responsible for Labs and Project)**
  - Dongding Lin, 22037064r@connect.polyu.hk, PQ719
  - Heming Xia, 23123186r@connect.polyu.hk, QT415

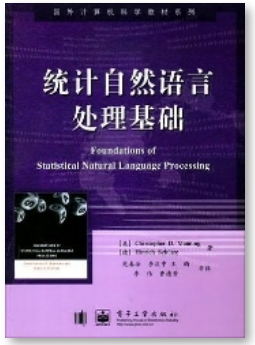


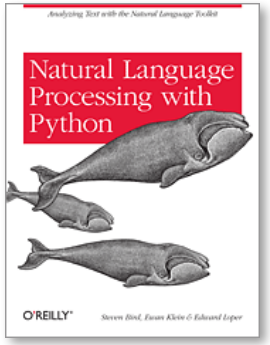
Course Introduction

- **Reference Books**
  - Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (P98.J87)
  - Foundations of Statistical Natural Language Processing (P98.5.S83 M36)





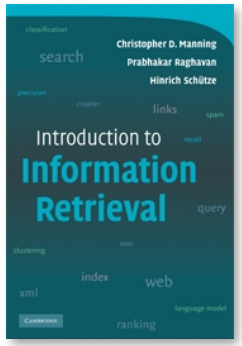


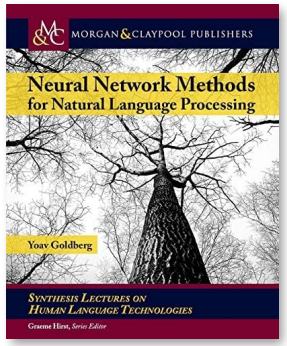


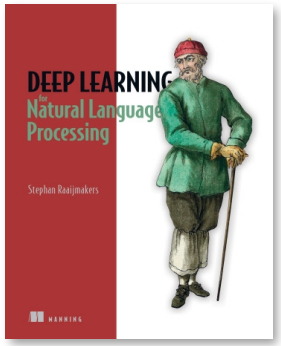
  - Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit

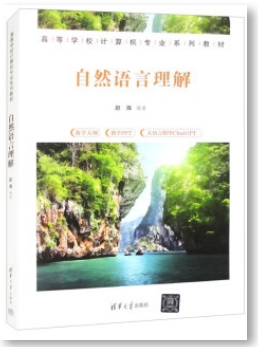
Course Introduction

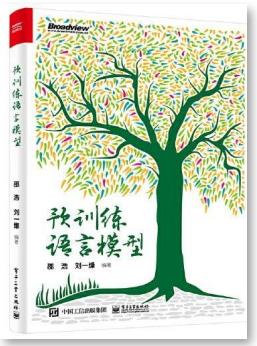
- **Reference Books**
  - Introduction to Information Retrieval (QA76.9.T48)











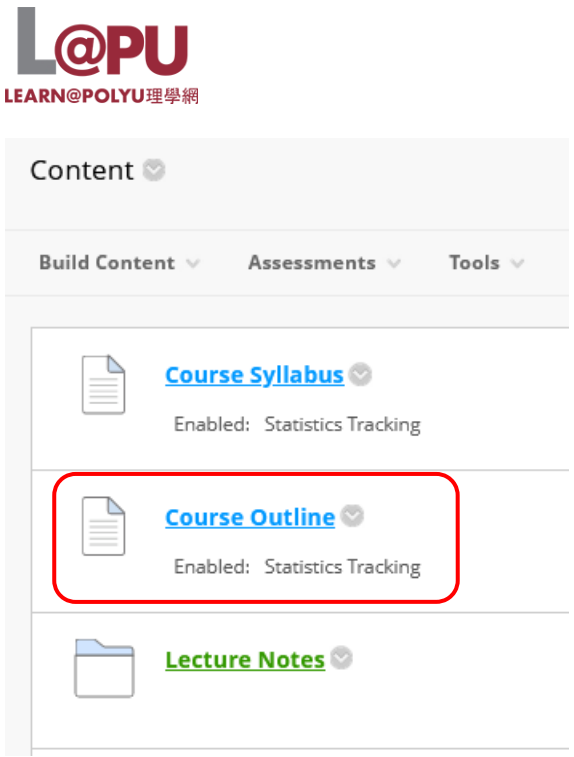
  - Neural Network Methods in Natural Language Processing (PolyU Library Online Access)
  - Deep Learning for Natural Language Processing (PolyU Library Online Access)

Course Introduction

- **Teaching Materials**
  - Lecture notes, lab materials, lecture homework, lab assignment, project description, and supplementary reading materials, etc. are downloadable from Blackboard.
  - Lecture notes and lab materials are normally available one day before class.
  - Submit your homework, assignment, project code and report to the Blackboard by due time. The penalty for late submission (no later than one week) is a reduction by up to 10% of the total mark.

Course Introduction

- **Tentative Teaching Arrangement and Schedule**
  - Lecture Sessions (FJ304)
    - Weeks: 1, 2, 3, 5, 6, 9, 11, 12 and 13
  - Lab Sessions (PQ604A/B/C)
    - Weeks: 4, 7 and 10
  - Mid-Term Quiz
    - Week: 8
  - See “Course Outline” in Blackboard for reference.

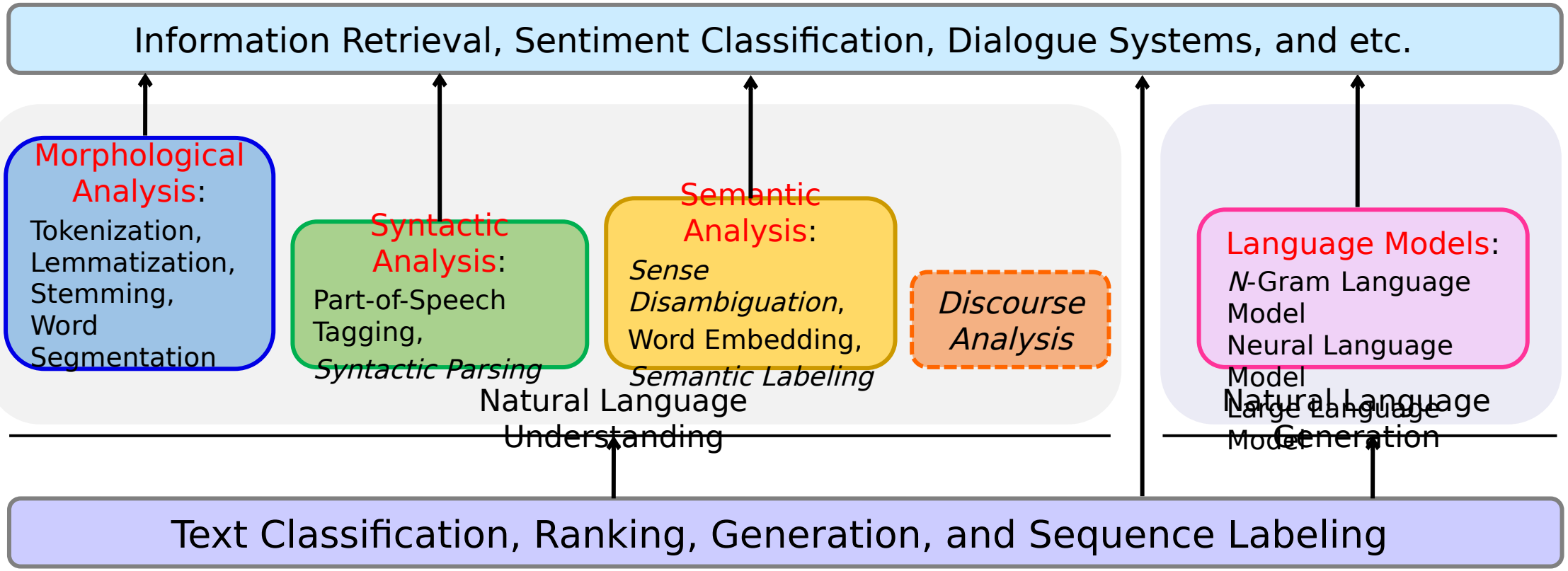


Course Introduction

- **Method of Assessment**
  - Course Work **55%**
    - Homework 5%
    - Lab Assignment 5%
    - Mid-Term Quiz 25%
    - Group Project 20%
  - Final Examination 45%

Course Introduction

- **Tentative Teaching Content**



Course Introduction

- **Tentative Teaching Content**
  - Lecture Sessions: NLP Concepts, Models, Approaches and Applications
  - Natural Language Understanding (NLU)
    - Morphological/Lexical Analysis: Tokenization, Stemming/Lemmatization, Chinese Word Segmentation
    - Syntactic Analysis: Part-of-Speech (POS) Tagging, Syntactic Parsing (Constituency Parsing, Dependency Parsing), Chunking, Context-Free Grammar (aka. Phrase-Structure Grammar)

Course Introduction

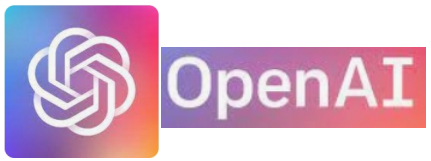
- **Tentative Teaching Content**
  - Natural Language Understanding (NLU)
    - Semantic Analysis: Word Sense Disambiguation, Word Embedding, Semantic Role Labeling
    - Discourse Analysis: Co-Reference Resolution, Coherence Modeling
  - Natural Language Generation (NLG)
    - *N*-Gram Language Models, Neural Language Models, Word-by-Word Generation, Rule-based Generation, Template-based Generation

Course Introduction

- **Tentative Teaching Content**
  - NLP Applications
    - Information Retrieval (Ranking)
    - Sentiment/Emotion Analysis (Classification)
    - Information Extraction and Knowledge Discovery (Sequence Labeling)
    - Question Answering (Classification, Information Retrieval and Information Extraction)
    - Machine Translation (NLU and NLG)
    - Text Summarization (Ranking, NLU and NLG)
    - Dialogue Systems (NLU, NLG, QA, IR, Recommendation)

Course Introduction

- **Tentative Teaching Content**
  - Lab Sessions: NLP Toolkits like NLTK and Stanford CoreNLP, Neural Language Models like BERT, GPTs and ChatGPT, Large Language Models like LLAMA, MISTRAL and etc.



  - Group Project: TBC

Course Introduction

- **Pre-Requisites**
  - No Pre-Requisites? Definitely NOT!
  - Machine Learning Knowledge: Classification (e.g., SVM, NB, Random Forest), Sequence Modeling (e.g., Hidden Markov Models), Neural Network Models (e.g., Feedforward Neural Networks)
  - Programming Skill: Python
  - Please find the "Supplementary Reading Materials" in Blackboard for your reference.

NLTK: Natural Language Processing with Python

Toolkit Download: https://www.nltk.org/

Book Website: https://www.nltk.org/book/

Youtube:

- Tokenizing Words and Sentences: https://www.youtube.com/watch?v=FLZvOKSCkxY&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=1
- Stop Words: https://www.youtube.com/watch?v=w36-U-ccajM&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=2
- Stemming: https://www.youtube.com/watch?v=yGKTphqxR9Q&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=3
- Part of Speech Tagging: https://www.youtube.com/watch?v=6j6M2MtEqi8&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=4
- Chunking: https://www.youtube.com/watch?v=imPpT2Qo2sk&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=5
- Named Entity Recognition: https://www.youtube.com/watch?v=LFXsG7fueyk&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=7
- Lemmztizing: https://learn.polyu.edu.hk/webapps/blackboard/execute/manageCourseItem?content_id=_3933048_1&course_id=_82088_1&dispatch=edit
- WordNet: https://www.youtube.com/watch?v=T68P5-8tM-Y&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=10
- Text Classification: https://www.youtube.com/watch?v=zi16nI82AMA&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=11
- Words as Features for Learning: https://www.youtube.com/watch?v=-vVskDsHcVc&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=12
- Naive Bayes: https://www.youtube.com/watch?v=rlSOsUaTrO4&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=13
- Scikit-Learn Incorporation: https://www.youtube.com/watch?v=nla4C-VYNEU&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=15

Stanford CoreNLP Toolkit

Download: https://stanfordnlp.github.io/CoreNLP/

Scikit-Learn: Machine Learning in Python

Website: https://scikit-learn.org/stable/

Python Tutorial

Enabled: Statistics Tracking

PyTorch Tutorials

Website: https://pytorch.org/tutorials/

TensorFlow Tutorials

Website: https://www.tensorflow.org/tutorials/index.html


thank you