CS310: Natural Language Processing

Instructor: Tom Ko

Information about the instructor

- Dr. Tom Ko (Assistant Professor in CSE)

- Office: Room 506, iPark A10 創園10棟

- Email: kot@sustc.edu.cn

- Office hour: by email appointment

Course objectives

- To learn pattern recognition

- To learn basic machine learning techniques

- To learn automatic speech recognition

- To learn natural language processing related knowledge

Grading scheme

- Lecture attendance (10%)
- Lab attendance (10%)
- Assignments and Project (50%)
- Final exam (30%)

Project

- Paper reading and presentation

Schedule

- 15 weeks => 15 lectures and 15 labs
- The lecture notes and lab outlines will be uploaded at the beginning of every weeks.
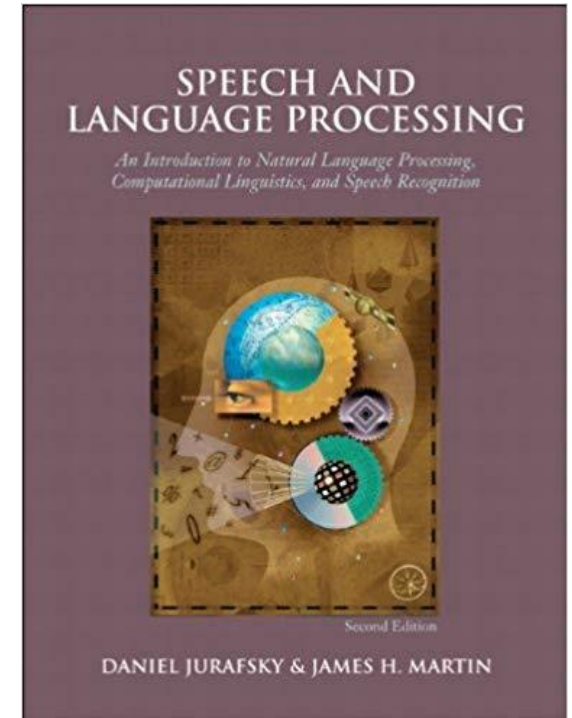- The description of the project will be announced at the middle of the course.

Textbooks

- ### Main textbook:

    - Speech and Language Processing, Daniel Jurafsky and James H. Martin, 2nd edition 2008.

- ### Reference books:

    - Richard O. Duda. **Pattern Classification**, 2nd edition, 2000.

    - *Spoken Language Processing: A Guide to Theory*, Algorithm, and System Development, 2001

Course site

- Our course site is at Sakai
  https://sakai.sustech.edu.cn

- Lecture notes, lab outlines, assignments can be downloaded from the course site.

Syllabus

- 70–80% Speech–related techniques
- 20–30% Language Processing

Syllabus

- 1) Introduction to ASR and NLP
- 2) Bayesian Decision Theory
- 3) Signal Processing in Speech
- 4) Hidden Markov Model and Gaussian Mixture Models
- 5) Deep Neural Network
- 6) Data augmentation skills
- 7) Speaker Verification
- 8) Regular Expression
- 9) Word2Vec
- 10) Text Classification
- 11) Machine Translation

Main aim of this course

‣ Not to classify you

‣ Hope you find the knowledge of this course useful in your life