Sungkyunkwan University (SKKU) International Summer Semester (ISS) 2024

Cultural Contents and Natural Language Processing
Prof. Youngeun Koo, Sungkyunkwan University

SHORT COURSE DESCRIPTION

This course focuses on Natural Language Processing(NLP) methods in cultural contents industry, such as movie, drama, music, and webtoon industry. This course aims to understand the importance of linguistic data analysis in developing cultural content services and learn basic techniques of natural language processing. Students will attempt to understand human(user/consumer) by analyzing their linguistic data and then develop human-centered services, such as personalized recommendation systems and culture-specific marketing strategies.

In this course, students will get a chance to practice basic Python grammars and NLP tools in class. The class will be conducted on the assumption that all students will have little (or no) prior knowledge of linguistics and programming.

READING MATERIALS

No reading materials. Course will be based on lecturer's slides.

COURSE REQUIREMENTS AND GRADING

Class **attendance** is essential for successful completion of the course, including assignments, mid-term exam and final project. Students are expected to attend every class session. If you are unable to attend any class session, please contact the lecturer beforehand. Since we will be practicing basic Python grammar and natural language processing tools during class, **students are recommended to bring their own laptops.**

Several **assignments** will be assigned for students to review basic grammars of python and tools for natural language processing, which are covered in the class. Assignments shall be done using Google Colab notebook. How to use Colab will be covered in class.

Mid-term exam is a written, short essay test, which ask whether students understand the concepts taught in class. For final exam, students will carry out a team project to design or build a cultural service based on natural language processing technique, and make a **final presentation** on it. Students are free to choose their team(max. 3 people).

All ISS classes are pass/fail based on the student academic achievement evaluated by grades on a scale of 100 points (grade of 60 or above is Pass). Grading policy is specified below:
  - Attendance: 15% (SKKU regulations require students to attend at least 80% of all classes.)
  - Assignments: 25%
  - Mid-term exam: 30%
  - Final presentation: 30%

COURSE SCHEDULE *(Schedule is tentative depending on the students' background knowledge!)*

– WEEK I –

Monday (1 July)
       What is Natural Language Processing(NLP)

Tuesday (2 July)
       Python Basics (1)
       Assignment 1: Data types and their functions

Wednesday (3 July)
      Python Basics (2)
      Assignment 2: Conditions, Loops, Class

Thursday (4 July)
      Data Crawling (1)

– WEEK II –

Monday (8 July)
      Data Crawling (2)

Tuesday (9 July)
      Data Preprocessing - Tokenization, normalization, cleaning
      Assignmet 3: Data crawling and preprocessing

Wednesday (10 July)
      Linguistic analysis (1) - Morphological analysis

Thursday (11 July)
      Linguistic analysis (2) - Syntactic analysis

– WEEK III –

Monday (15 July)
      Linguistic analysis (3) – Semantic analysis

Tuesday (16 July)
      Mid-term exam

Wednesday (17 July)
      Machine Learning & Deep Learning based NLP

Thursday (18 July)
      NLP in Movie/drama Industry

– WEEK IV–

Monday (22 July)
      NLP in Music/K-pop Industry

Tuesday (23 July)
      NLP in Game and Webtoon Industry

Wednesday (24 July)
      Final presentation
      - Presentation on team project