

특강 계획서

언어학 기반의 자연어처리

담당강사	성명		문성민	
	소속	아주대학교 인문과학 연구소 연구조교수		
	e-mail	stat34@ajou.ac.kr	홈페이지	http://seongminmun.com/

1. 특강 목표 및 개요

<p>본 특강은 '디지털 인문학 (Digital Humanities)'이라는 큰 테두리 안에서 언어학 기반의 자연어처리 방법을 소개하고 Python을 사용하여 실습을 진행하는데 목적을 두고 있다. 구글에서 제공하는 Colab을 사용한 본 특강은 코드를 설계하고 작성하는 기술에 초점을 두기 보다는 실제 데이터를 언어학 기반으로 분석하는 방법에 초점을 둬므로써 특강을 수강하는 연구자들로 하여금 실질적으로 연구에 도움이 되는 방법을 제공하는데 목적을 두고 진행될 예정이다.</p> <p>특강은 총 4회에 걸쳐 진행될 예정이며 각 회차마다 강의자가 진행하였던 각기 다른 자연어 처리 기반의 양적 언어 연구들을 간략히 소개하고 실제 분석 과정을 실습하는 형태로 진행 될 것이다. 또한 실습이후에는 토론을 통해 특강에 참여하는 연구자들로 하여금 실습한 내용을 자신의 것으로 습득할 수 있는 시간을 가질 것이며, 학술적인 연구 토론을 통해 연구자들로 하여금 새로운 연구 주제를 모색하는 시간이 될 예정이다.</p>
--

2. 각 회차 별 특강 계획

주별	강의 내용	강의 형태	세부 강의 내용
6/12	음운론적 발음 변환 규칙을 사용한 발음변환기	이론, 실습	음운론적 발음 변환 규칙과 정규 표현식을 사용하여 기저형의 문장을 표면형로 변환하는 발음 변환기(i.e., G2P)에 대한 이론을 학습하고 파이썬을 통해 구현한다.
6/19	성별에 따라 이름에서 나타나는 음성상징성 (i.e., Sound symbolism)	이론, 실습	성별에 따라서 이름에서 나타나는 음운론적 특성을 분석하고 로지스틱 회귀 분석을 사용하여 이름을 통해 성별을 예측하는 모델을 구현한다
6/26	단어 임베딩 모델을 사용한 문학 작품간 유사성 분석	이론, 실습	비지도 학습 방법으로 TF-IDF와 같은 단어 임베딩 모델의 이론을 학습하고 문학 작품간 유사성을 기반으로 네트워크를 생성하는 알고리즘을 파이썬을 통해 구현한다.
7/3	언어 모델(i.e., LLM)을 사용한 혐오표현탐지 (i.e., Hate speech detection)	이론, 실습	BERT와 GPT와 같은 언어 모델(i.e., LLM)을 학습하고 문장에서 드러나는 혐오표현을 탐지하는 알고리즘을 파이썬을 통해 구현한다.