

수업 계획서

[방학 특강] R 데이터(텍스트) 마이닝 및 시각화 분석

강의	강의 시간	20시간		
	주수강대상 학부/전공/학년	아주대학교 사학과	개설년도/학기	2021/겨울방학
	강의일정 및 강의실	2022년 1월 17일 - 28일 / 온라인 수업		

담당 강사	성명(직위/소속)	문성민		
	강사 소개	파리 낭테르 대학교에서 전산언어학으로 박사학위를 수여 받았으며 주요 연구 분야는 자연어 처리, 데이터 시각화, 기계학습이다.		
	e-mail	simon.seongmin.mun@gmail.com		
	홈페이지	http://seongminmun.com/		

1. 강의 요약

프로그래밍 활용 능력에 대한 수요에 맞추어 프로그래밍 언어 중 사용 빈도가 높고 데이터 분석에 적합한 R 언어를 학습한다. 또한 교과과정을 통해 수강생들에게 두가지 학습성과를 제공한다.

첫째, R 언어의 사용 방법과 활용 방법을 자세히 학습하고 이해함으로써 프로그래밍에 대한 두려움을 감소시키고 자신감을 향상시킨다.

둘째, R로 자연어 처리를 학습함으로써 전산언어학 분야를 소개하고 실제 역사 데이터를 활용하여 텍스트 마이닝을 하는 작업 과정을 파악한다.

상기의 학습성과를 갖추게 되면, R을 활용하여 업무를 수행하는 분야(i.e., 금융, 통계조사, 연구소, etc.)로 진로를 나아가고자 하는 학생들은 현업에서 실제로 필요한 기술을 습득하게 될 것이며, 실제 업무와 비슷한 커리큘럼의 학습을 통해 취업 과정에서 필요한 경험을 하게 될 것이다.

2. 강의 목표

수강생들에게 다소 어렵게 느껴지는 프로그래밍을 차근차근히 배움으로써 프로그래밍 활용에 대한 두려움을 낮추고 자신감을 향상시킨다. 또한 자연어 처리의 전반적인 과정 (i.e., 데이터 수집, 자연어 처리, 통계분석, 시각화)을 체계적으로 배움으로써 자연어 처리에 대한 이해를 높이고 프로그래밍을 활용한 자연어 처리 기술을 학습한다.

수업 계획서

3. 강의 진도

일별	시간	강의 목표	세부 학습 내용
17 일(월)	17 시-19 시 (2 시간)	R 기본 다지기 1	R 설치하기 자료형 알아보기 (벡터, 리스트, 행렬, 데이터 프레임)
18 일(화)	17 시-19 시 (2 시간)	R 기본 다지기 2	제어문 & 조건문 사용하기 함수 만들기
19 일(수)	17 시-19 시 (2 시간)	데이터 수집 1	정적 웹사이트 크롤링 (네이버 영화)
20 일(목)	17 시-19 시 (2 시간)	데이터 수집 2	동적 웹사이트 크롤링 (조선왕조실록)
21 일(금)	17 시-19 시 (2 시간)	데이터 정제	KoNLP 사용하기 (형태소 분석) dplyr 을 이용한 데이터 변환
24 일(월)	17 시-19 시 (2 시간)	데이터 분석 1	기초통계(평균, 분산, 표준편차) 카이 동질성 검정 Two sample t-test
25 일(화)	17 시-19 시 (2 시간)	데이터 분석 2	분산 분석 (ANOVA) 회귀 분석 (Regression)
26 일(수)	17 시-19 시 (2 시간)	데이터 시각화 1	ggplot2 를 이용한 시각화
27 일(목)	17 시-19 시 (2 시간)	심화 과정 1 (조선 왕들의 성향 파악)	데이터 수집 및 정제 (조선왕조실록)
28 일(금)	17 시-19 시 (2 시간)	심화 과정 2 (조선 왕들의 성향 파악)	데이터 분석 및 시각화 (군집분석 & t-SNE 시각화)