

KDI 국제정책대학원

2019년 3분기 연구방법론세미나 강의계획서

□ 텍스트자료분석

1. 과정개요

담당강사	성명	E-mail
	윤호영	hoyoungemail@gmail.com
강의시간 / 강의실	8월 17일~ 9월 25일 / 온라인 E-education 페이지(추후 공지)	
강의목표 및 개요	본 수업은 R을 활용한 텍스트 분석을 실시하기 위한 강좌로, 텍스트 분석과 관련된 기본 이론과 실습 코드를 통한 실전 분석을 수행하는 과정이다. 분석과 관련해서는 단어 빈도 분석, 토픽 모델링(LDA 및 STM), 단어 임베딩, Support Vector Machine을 활용한 문서 분류의 4가지를 강의한다. 실시간 강의는 네이버 뉴스 제목 스크래핑과 Q&A를 진행할 예정.	
참가 대상	학부, 대학원생 및 일반 연구자 (사회과학에서 사용하는 통계적 연구방법에 관심 있는 사람들은 누구나)	
선수 과목	없음. 프로그램 설치와 같은 기초적인 사항으로부터 강의가 시작될 예정. 초급에서 중급까지의 과정임.	
수업운영방식	온라인 강의 기반이며 과제를 부여합니다.	
교재 및 참고문헌	강사의 PPT 노트와 다음의 내용 원고가 제공됩니다. 윤호영(2019). 제 2장 비정형 텍스트 데이터 분석: 통계적 언어처리 - 토픽모델링. “빅데이터의 분석방법과 활용”, 정원준 외, 광고 지성총서, 제 9 권. 학지사 참고문헌으로는 “R로 배우는 텍스트 마이닝 : tidytext 라이브러리를 활용하는 방법” 책이 있습니다.	
실습자료	강사가 제공하는 실습용 자료 (추후 공지)	
사용프로그램	R	
기타 사항	(강사님 코멘트) R, R Studio 설치가 필요합니다. 단, 현재 형태소 분석 패키지인 KoNLP의 호환성 문제로 R 4.0이 아닌 3.6 버전이 필요합니다.	
담당교수 정보	[학력] 2019 방송통신대학교 컴퓨터과학 학사 2017 위스콘신대 언론학 박사 2010 일리노이대 사회학과 박사과정 3년 마침. 2005 연세대학교 사회학 석사 2002 연세대학교 영어영문학 학사 [수상경력] 정보사회학회 신진연구자 우수연구상 사이버커뮤니케이션학회 신진연구자 우수상 Dissertator Fellowship The Louise Elizabeth George Scholarship Fund Journalism & Mass Communication, UW-Madison	

	[자격증] 데이터분석 전문가[국가공인], 한국데이터산업진흥원 SQL 개발자[국가공인], 한국데이터산업진흥원 데이터 아키텍처 준전문가, 한국데이터산업진흥원 정보처리기사[국가기술자격증], 과학정보통신부 [최근 3년간 주요 저서/논문] 1. Ahn, S. J., Yoon, H. Y.. (In-Press). Green Chasm' in clean-tech for air pollution: patent evidence of a long innovation cycle and a technological level gap. Journal of Cleaner Production. 2. 윤호영(2019). 제 2 장 비정형 텍스트 데이터 분석: 통계적 언어처리 - 토픽모델링. "빅데이터의 분석방법과 활용", 정원준 외, 광고 지성총서, 제 9 권. 학지사 3. 윤호영(2019). 제 6 장 빅데이터 분석 알고리즘과 커뮤니케이션: 데이터 테크놀로지 시대의 분석 전략의 변화. 광규태 외. "데이터 테크놀로지와 커뮤니케이션 연구". 커뮤니케이션북스 4. 윤호영 (2019) 빅데이터 시대의 데이터 저널리즘 교육에 관한 소고: 무엇을 어떻게 할 것인가? 한국언론학보, 87-115 5. 윤호영 (2019) 팟캐스트 방송의 생태계의 진화: 다양성 변화, 경쟁 및 생존 한국방송학보, 33(4), 50-98 6. 윤호영 (2018) 기술-사회 패러다임 용어는 어떻게 소비되는가? 탐색적 접근 : "4차 산업혁명"에 관한 26만건 뉴스(제목) 분석, 정보사회와 미디어, 19(2), 1-28 7. Kim, K., Yoon, H. Y., & Jung, K. (2017). Resilience in Risk Communication Networks: Following the 2015 MERS Response in South Korea. Journal of Contingencies and Crisis Management, 25(3),148-159. 8. Lee, M., Yoon, H. Y., & Park, H. W. (2017). From online via oine to online: How online visibility of tourism information shapes and is shaped by offline visits. Journal of Travel & Tourism Marketing, 34(9), 1143-1154.
--	--

2. 강의계획

일시	강의 및 실습 내용
1강	R, RStudio 프로그램 설치, 텍스트 분석 강의 내용과 관련된 R 기초
2강	텍스트 분석의 이론적 이해
3강	단어 빈도 분석 실습과 해석 방법
4강	벡터 공간 모델: 단어 임베딩의 원리 및 GloVe 분석을 통한 키워드 해석 충실화
5강	토픽모델링 : LDA, STM, CTM
6강	기계학습의 기초 원리 및 Support Vector Machine을 활용한 텍스트 분류
7강 - 8강	실시간 과제 (스크래핑) 및 Q&A