

<독일학 연구 및 교육을 위한 디지털 인문학: 기초>

디지털인문학 연구를 위한 프로그래밍언어 Python 기초와 활용: Google Books Ngram Viewer 데이터의 검색 및 분석

홍정하 (서강대 인문콘텐츠 융합소프트웨어 연계전공 연구교수)

프로그래밍언어 Python은 데이터의 수집·가공·분석·시각화에 용이한 환경을 갖추고 있어 데이터과학(data science) 및 디지털인문학에서 프로그래밍언어 R과 더불어 활용도가 높은 프로그래밍언어이다. 본 특강에서는 데이터 분석에 효과적인 프로그래밍언어 Python을 이용한 Google Books Ngram Viewer (<https://books.google.com/ngrams>) 데이터를 검색·가공·분석 방법을 소개하고 이를 통해 빅데이터를 이용한 디지털인문학 연구의 방법과 가치를 탐색한다. 본 특강에서 다루는 Google Books Ngram Viewer 데이터는 전 세계 유명 도서관에서 소장하고 있는 영어(미국, 영국, 소설), 중국어(간체), 프랑스어, 독일어, 히브리어, 이탈리아어, 러시아어, 스페인어 문헌 및 잡지로 구성된 대규모 코퍼스로부터 어휘 및 어휘 연쇄(n-gram)의 빈도 분포 추세를 1800년부터(일부 데이터는 1500년부터) 2019년까지 검색할 수 있도록 Google이 개발하여 공개한 데이터이다. 「Michel, Jean-Baptiste, et al. "Quantitative analysis of culture using millions of digitized books." *science* 331.6014 (2011): 176-182.」에서는 Google Books Ngram Viewer 데이터를 이용한 인문사회과학 연구 사례를 제시하고 있으며, 본 특강에서는 Michel et al. (2011)의 연구 사례 중 미국과 독일의 “정치적 검열과 탄압” 사례를 중심으로 Google Books Ngram Viewer 데이터를 검색하여 검색 결과를 웹스크래핑(web scraping)하는 방법부터 데이터 분석을 위한 기본적인 데이터 가공 및 시각화 자료 생성 방법을 다룬다. 특히, 프로그래밍언어 학습 경험이 없는 수강생을 고려해 1회차 및 2회차 강의는 Python 설치법 및 최소한의 필수 개념과 기본 사용법을 다룬다.

－ 일시: 7월12일 ~ 8월 16일 (매주 월요일) 14:30 ~ 16:30

－ 장소: 서관 316A

* 참여 예상인원: 약 10명 * 준비물: 노트북

－ 프로그램 －

일시	주제
7월 12일(월)	Python 및 디지털인문학 소개 Python 설치 및 기본 사용법
7월 19일(월)	Python 기초 문법과 기본 데이터구조
7월 26일(월)	Python을 이용한 검색/웹스크래핑 및 데이터 가공
8월 02일(월)	Python을 이용한 차트 만들기
8월 09일(월)	Python을 이용한 디지털인문학 연구 사례 I
8월 16일(월)	Python을 이용한 디지털인문학 연구 사례 II

주최: 고려대학교 독어독문학과

주관: 독일어권문화연구소

후원: 고려대학교 대학원