

역사 텍스트 디지털화 입문

OCR에서 데이터베이스까지

이미지 > OCR 추출 > 검수 > 구조화 > 데이터베이스

일시	2026년 3월 3일(화) 오후 8:00 ~ 10:00
형식	줌(Zoom) 비대면 / 2시간
강사	허 수 (서울대학교 역사학부)
대상	인문학 연구자 (디지털 도구 입문자 환영)

강의 내용

전반부 (50분)

이미지에서 텍스트 꺼내기

- 라이브 시연: AI로 100년 전 책 읽기
- 실습 1: 어린이 잡지 OCR
- 실습 2: 신인철학 OCR
- 시연: 일본어 잡지 목차 OCR
- OCR 결과 검수 방법

후반부 (50분)

텍스트를 데이터베이스로 만들기

- 완성 DB 시연 (2,254행 실물)
- 구조화: 문장 단위 DB 개념
- 실습 3: AI로 TSV 구조화
- 실습 4: Google Sheets DB 구축
- 다음 단계 안내 + Q&A

사용 도구 및 준비물

- 도구: Google AI Studio (aistudio.google.com) — 무료, 별도 설치 없음
- 준비물: Google 계정 + PC (듀얼 모니터 또는 큰 화면 권장)

참여 안내

줌 접속: <https://snu-ac-kr.zoom.us/j/82898609376?pwd=Aeqj4E3wOiNXvGWbaCT8QE15mfJTmO.1>

실습 자료: https://drive.google.com/drive/folders/1Dt9AZHiNs5MPKURMd00v5ELhkuW_3Wt

문의: 이혜경 (yhgyung@snu.ac.kr) | 허수 (crctaper@snu.ac.kr)