

(첨부 1)

-2025년 한국간호교육학회 하계 교수법 학술워크숍-
주제: ZERO-coding 자연어 처리 교육 및 실습 : KNIME 활용

1. 교육 개요

- 교육 배경: 자연어처리는 다양한 분야에서 활용도가 높아지고 있으며, 이를 쉽게 접근할 수 있는 오픈소스, 노코딩 툴로 KNIME이 각광받고 있음
- 교육 목적: KNIME을 활용한 실습 중심의 자연어 처리 교육을 통해 참여자의 연구 역량 강화

2. 교육 개요

- 주최: 한국간호교육학회
- 주제: ZERO-coding 자연어 처리 교육 및 실습 : KNIME 활용
- 일정: 2025년 8월 6일(수) 09:00~16:00 (점심시간 1시간 포함)
- 대상: 자연어 처리에 관심 있는 교수, 대학원생, 대학생, 실무자
- 인원: 20명 내외
- 장소: 온라인

3. 교육 내용 및 커리큘럼

- KNIME 소개 및 설치
- 텍스트 전처리 실습
- 워드클라우드 및 시각화
- LDA 토픽 모델링 실습

4. 강사 프로필

- 서울대학교 간호대학 우경미 교수, 서울대학교 간호과학연구소 김애리 선임연구원
 - 자연어 데이터 분석 프로젝트, 논문 출판 다수 수행
 - KNIME, Python 기반 자연어 처리 경험 다수

5. 교육생 사전 준비 사항

- 개인 소장 데스크탑 혹은 노트북에 KNIME 설치
- 강의 참석을 위한 비디오, 마이크와 실습 환경: KNIME 설치 안내 동영상 참조 (Window용, MacOS용)

<강의안>

Time	Contents
09:00~09:50	* 환영사, 강사소개, 단체사진 촬영 * 오전은 이론으로 구성 (1) Exploring Text Mining in Big Data - Leveraging Healthcare Unstructured Data a. 빅데이터의 정의 -빅데이터란? -의료분야의 다양한 빅데이터 b. 의료에서의 정형, 비정형 데이터 및 분석법 -정형데이터와 비정형데이터 -텍스트마이닝 -자연어분석법 c. 기계학습 -Machine learning (ML) 정의 -ML 연구 현황 및 트렌드 -ML 종류: supervised vs unsupervised 장단점 d. 의료에서 빅데이터를 활용한 텍스트 마이닝 연구 -의료 데이터를 활용한 텍스트 마이닝 및 기계학습 활용
10:00~10:50	(2) Topic modeling과 KNIME a. Topic modeling -필요성 -토픽모델링이란? -Background -Fundamental Concepts in Topic Modeling -LDA토픽모델링 개념 -LDA 토픽모델링 도식화 -LDA 토픽모델링 이해

-LDA 토픽모델링의 예시

-LDA 토픽모델링 연구 적용 프로세스 예시

b. KNIME을 활용한 토픽모델링

-KNIME이란?

-KNIME의 구성

-KNIME의 활용

참고자료: (부록) 토픽모델링을 활용한 해외연구사례

11:00~11:50

(3) 자연어 분석의 개념설명 (Concept summary)

a. 자연어분석법의 과정

-문장분할(sentence segmentation)

-토큰화(tokenization)

-어간추출(stemming)

-표제어추출(lemmatization)

-불용어처리(stop word)

-의존 구문분석(dependency parsing)

-품사태깅(Part of speech, POS tagging)

b. TF-IDF(Term Frequency-Inverse Document Frequency)

c. 최적의 토픽수 구하기

12:00~13:00

점심시간

13:00~13:50

* 오후는 실습으로 구성

(4) KNIME Workflow와 node사용하기

- Workflow 란? Node 란?

- Workflow 만드는 기본 방법

- Workflow 검색 방법 및 다운로드, 적용, 반입 및 반출

- Node에 대한 기본적인 설명

14:00~14:50

(5) 텍스트 데이터 수집하기

- 검색어 구축하기

-문서 추출하기

- Histogram 만들기

(6) 텍스트 정제하기: Preprocessing

- 데이터 전처리

15:00~15:50

(7) 핵심어 빈도 분석 (Frequency vs TF-IDF)

- 단순 빈도분석

- TF-IDF

- 워드클라우드 만들기

(8) LDA 토픽 모델링 I: 최적의 토픽 수 찾아내기

- Perplexity*

- Elbow method

- 최적의 토픽 수 결정

(9) LDA 토픽 모델링 II

- Topic Extractor node

- 토픽별 핵심어와 워드클라우드 만들기

Q&A

폐회사, 설문지 URL 송부

*는 쉬는 시간 (노드 수행 시간이 오래 걸림)