

**강좌 키워드	언어 데이터, 텍스트 데이터, 비정형 데이터, 텍스트 처리, 자연언어처리								
*1. 수업 목표	자연어 데이터 분석을 위해 딥러닝 모형을 비롯한 최신의 자연어처리 기법을 학습할 필요가 있으나 컴퓨터 언어학이나 자연어처리 관련 배경 지식이 없는 다양한 분야의 전공자들이 바로 접근하기에는 어려움이 있다. 특히 컴퓨팅과 데이터 처리의 경험이 많지 않은 전공자들에게는 자연어처리 기술을 학습하기에 앞서 언어 데이터 처리를 위한 컴퓨터 활용의 기초를 학습하는 것이 도움이 될 것이다. 정형 데이터는 이미 형식을 갖추고 있어 분석과 모델링 단계로 수월하게 진행이 가능하지만 비정형 데이터는 데이터 자체를 다루기 위한 처리에 많은 노력을 요구한다. 이러한 배경에서 '언어 데이터 처리'를 위한 컴퓨팅 기초를 학습할 수 있도록 한다.								
**2. 교재 및 참고문헌	교재	PDF 자료 제공							
	참고문헌	Janssens, J. (2014). Data Science at the Command Line: Facing the Future with Time-Test Tools. O'Reilly.							
**3. 강의계획	주요 수업방식	■플립러닝 □이론 위주 수업 □토론 위주 수업 □프로젝트 수업 □기타							
	<p>이 과목은 텍스트로 된 언어 데이터를 정리하고 처리하기 위한 컴퓨팅 개념 및 활용의 기초 교과목이다. 빅데이터 언어인지 분야를 공부하려는 인문계열 초급 학습자를 대상으로 하며 자연어처리에 관련된 선행 지식은 요구하지 않다. 빅데이터의 대부분은 비정형 데이터로 이루어져 있으며 그 전형적 대표인 자연어 텍스트 데이터는 많은 시간의 집중적인 데이터 수집, 정제, 전처리를 거쳐야만 데이터를 이해하고 심화된 분석으로 나아갈 수 있다. 이러한 맥락에서 이 수업은 수강생들이 디지털 데이터를 이해하기 위한 배경 지식을 쌓고 적절한 기법과 도구를 활용하여 텍스트 데이터를 처리하여 더 정교한 자연어처리를 위한 준비를 할 수 있도록 하는 데에 목표를 두고 있다. 수업의 주제로는 파일, 디렉토리, 명령행 인터페이스, 텍스트 에디터와 같은 기본적인 컴퓨팅 주제부터 시작하여 문자셋, 인코딩, 마크업 언어, 데이터 파일 형식, 주석, 텍스트 처리 도구 등의 주제가 포함될 것이다.</p> <p>1 : 강의소개  2 : 소프트웨어 도구: 컴퓨팅 환경 준비하기  3 : 문자코드와 인코딩 (1)  4 : 문자코드와 인코딩 (2)  5 : 데이터와 텍스트  6 : 마크업  7 : 개방형 표준, 테이블형 데이터  8 : 파일과 디렉토리; 텍스트 처리 도구 기초 (1)  9 : 텍스트 처리 도구 기초 (2); 검색  10 : 정렬; 빈도  11 : 속성-값 데이터  12 : 테이블형 데이터  13 : 텍스트 치환  14 : AWK 프로그래밍  15 : 텍스트 처리 실습; 질의응답</p>								
*4. 평가방법	성적부여 방식	절대평가							
	등급제 여부	S/U(성적평가방법(A~F, S/U) 변경 가능) ※S/U : 급락제의 서울대학교 표기법으로 P/NP, P/F 등과 같은 의미임							
	구분	출석	과제	중간	기말	수시평가	태도	기타	합계
	비율	30%	70%						100%
	비고								
	출석 규정	수업일수의 1/3을 초과하여 결석하면 성적은 "F" 또는 "U"가 됨(학칙 85조) 결석에 대하여 교원에게 별도로 출석인정을 받은 경우 예외로 할 수 있음 (학업성적처리규정. 조기취업자 출석 및 성적처리 지침)							
	기타 사항	시험 없음. 문제풀이 및 보고서 과제.							
5. 정원 외 신청	추가 수용 인원	최대 60명							
6. 수강생 참고사항	선이수 교과목	없음							
	수강 시 필요사항	없음							

	면담시간 및 장소	ETL을 통해 사전 연락 요망
7. 장애학생 지원사항  ※ 필요에 따라 내용 수정 가능	강의수강 관련	○ 시각장애: 교재 제작(디지털교재, 점자교재, 확대교재 등), 대필도우미 허용 ○ 지체장애: 교재 제작(디지털교재), 대필도우미 및 수업보조 도우미 허용 ○ 청각장애: 대필 및 문자통역 도우미 활동 허용, 강의 녹취 허용 ○ 건강장애: 질병 등으로 인한 결석에 대한 출석 인정, 대필도우미 허용 ○ 학습장애: 대필도우미 허용 ○ 지적장애/자폐성장애: 대필도우미 및 수업 멘토 허용
	과제 및 평가 관련	○ 시각장애/지체장애/청각장애/건강장애/학습장애: 과제 제출기한 연장, 과제 제출 및 응답 방식의 조정, 평가 시간 연장, 평가 문항 제시 및 응답 방식의 조정, 별도 고사실 제공 ○ 지적장애/자폐성장애: 개별화 과제 제출 및 대체 평가 실시
	비고	본 강의를 수강하는 장애학생들에게는 이상의 지원 서비스 이외에도 장애학생 개개인의 특성과 요구에 따라, 지도교수 및 장애학생지원센터와의 상담을 통하여 적절한 수준의 지원 서비스를 제공합니다. 장애학생에 대한 지원서비스와 관련하여 문의사항이 있는 학생들은 담당교수 *** (02-880-****) 혹은 장애학생지원센터(02-880-8787)로 문의바랍니다.

- ◎ 강의계획서 직접입력 시 필수 입력 항목: \*, \*\*
- ◎ 강의계획서 첨부파일 업로드 시 시스템 필수 입력 항목: \*\*

**Course Keywords	language data, textual data, unstructured data, text processing, natural language processing								
*1. Goals	For natural language data analysis, it is necessary to learn the latest natural language processing techniques, including deep learning models, but it is difficult for those majoring in various fields without background knowledge in computer linguistics or natural language processing to directly access it. In particular, it will be helpful for majors who do not have much experience in computing and data processing to learn the basics of computer utilization for language data processing before learning natural language processing technology. Structured data already has a format, so it can be easily proceeded to the analysis and modeling stage, but unstructured data requires a lot of effort in processing to deal with the data itself. In this background, it is possible to learn the basics of computing for 'linguistic data processing'.								
**2. Reading Materials	Textbooks	PDF							
	References	Janssens, J. (2014). Data Science at the Command Line: Facing the Future with Time-Test Tools. O'Reilly.							
**3. Course Schedule	Lecture Method	<input checked="" type="checkbox"/> Flipped learning <input type="checkbox"/> Theory-driven <input type="checkbox"/> Discussion-oriented <input type="checkbox"/> Project-based <input type="checkbox"/> Others							
	Week 1: Software tools Week 2: Character code Week 3: Unicdoe Week 4: Hangul and encoding Week 5: Linux: Files and directories Week 6: Character code Week 7: Data Week 8: Text Week 9: Table data Week 10: Linux text processing 1 Week 11: Linux text processing 2 Week 12: Linux text processing 3 Week 13: Markup Week 14: Linux text processing 4 Week 15: Linux text processing 5								
*4. Evaluation	Grading Method		Absolute evaluation						
	Grading Type		S/U						
	Item	Attendance	Assignment	Midterm	Final	Quizzes	Attitude	Other	Total
	Rate	20	80						100%
	Note		5 times during the semester						
	Attendance Policy		Students who are absent more than 1/3 of class days will receive “F” or “U” grade.Students whose attendance is acknowledged can be exceptions. (Academic Grading Regulations, Guidance of Attendance and Grading for Early Employed Students)						
	Other		Other matters pertaining to the evaluation method such as regulations on cheating, whether and how alternative tests are made, and whether feedback for assignments or tests is provided						
5. Quota Exceeding Course Registration	Capacity	Up to 60 Students							
6. Guideline for Students	Prerequisite Courses								
	Requirements								
	Office Hours		by appointment						
7. Support Services for Students	For Lectures	<input type="radio"/> Visual Impairment: Make textbooks(digital textbook, braille textbook, enlarged textbook etc.), Allow note takers <input type="radio"/> Physical Disability: Make textbooks(digital textbook), Allow note takers and assistants <input type="radio"/> Hearing Impairment: Allow note takers and translators, Allow lecture recording <input type="radio"/> Health Impairment: Excuse absence due to health problems, Allow note takers <input type="radio"/> Learning Disability: Allow note takers <input type="radio"/> Intellectual Disability / Autism Spectrum Disorder: Allow note takers and							

- ⦿ Required fields when directly typing syllabus in mySNU: \*, \*\*
- ⦿ Required fields when uploading syllabus file in mySNU : \*\*