

팀 미팅 회의록

팀명	4726	차수	1 차
일 시	2019 년 2월 16일 토요일 19시 0분 - 22시 0분 (3시간 00분)		
장 소	성동구 행당동 스터디룸 '모임'		
참석자	고현경, 김희주, 김혜인, 이선희, 이수민		
불참자			
안 건	자기소개서 분석을 위한 자연어처리 단계는 어떻게 되는가?		
회의내용	<p>1. 개발 목표 확정: 자연어 처리를 활용하여 자기소개서를 분석한다.</p> <p>2. 자연어 처리 단계 정리:</p> <ol style="list-style-type: none"> 1) 기업별 합격 데이터를 수집한다. 2) 자소서를 분석하기 위해 형태소 분석을 한다. -> 형태소 분석은 KoInPy 라이브러리를 활용하여 정제하도록 한다. 3) 데이터 벡터화를 진행 word2vec을 통해 벡터화를 진행한다. 4) 벡터화된 데이터를 군집화를 하도록 한다. -> 군집화에 대한 자세한 연구가 더 필요하다. <p>자연어 처리를 위한 python 라이브러리 종류</p> <ul style="list-style-type: none"> • KoNLPy <ul style="list-style-type: none"> ○ 한국어 자연어처리를 위한 대표적인 파이썬 라이브러리 ○ Twitter, Komoran, Mecab 등 다양한 형태소 분석기 내장 • NLTK <ul style="list-style-type: none"> ○ 영어로 된 텍스트의 자연어처리를 위한 대표적인 파이썬 라이브러리 • Gensim <ul style="list-style-type: none"> ○ 주로 Topic modeling, Coupus(말뭉치) 및 word embedding 모델을 지원해줌 ○ 한국어 지원 가능 		