|  |  |
| --- | --- |
| **[2019-데이터크롤링] 기말 과제 계획서** | |
| **작 성 자** | **학과 : 컴퓨터공학과 / 학번 : 2017305057 / 이름 : 이성근** |
| **프로젝트 명** | The Reason |
| 프로젝트 목표 | Web Crawlring 을 활용해 수집하고, 비지도 학습과 Back Of Words 를 활용해 자기소개서를 Clustering 및 Text Minig 한다. 군집화된 자기소개서의 Topic Modeling 후 (모의면접)질문과의 키워드 맵핑 및 데이터 시각화를 목표로 한다. |
| 프로젝트 수행을 위한 학습내용 | - Data Crawling  - Text Mining  - 비지도 학습을 활용한 Clustering  - Topic Modeling  - Back Of Words  - Words To Vectors |
| 데이터 수집  계획 | 사람인, 자소설닷컴 등 각종 취업 정보사이트의 자기소개서를 Data Crawling을 활용해 수집한다. |
| 데이터 준비 및 탐색 작업 계획 | 수집한 데이터를 비지도 학습과 Back Of Words를 활용해 Clustering 후 Text Mining 및 Words To Vectors를 활용해 Topic Modeling 한다. |
| 분석 모델 구축  계획 | Clustering 된 자기소개서의 키워드 분석을 해 Topic Modeling을 한다. Topic Modeling된 자기소개서의 공통점과 서로 다른 군집간의 차이점을 분석한다. |
| 결과 시각화  계획 | Clustering 된 자기소개서의 Key Words 를 Words To Vector 를 활용해 Word Cloud 형식으로 시각화 하고, 기존의 한이음 프로젝트의 자기소개서 기반 모의면접 챗봇의 질문과 맵핑한다. |