|  |  |
| --- | --- |
| **Team Leader** | **임기찬** |
| **Team Member** | **강민수 김상헌**  **오준오 천영재** |
| **Professor** | **정재은** |

**CauNlpResume**



1. **과제 개요**

같은 뜻을 가진 단어라도 표현하는 바에 따라 받아들여지는 느낌이 다르다. 자기소개서와 같은 공적인 문서의 경우 이러한 경향이 더 커져서 같은 뜻을 가진 단어를 사용하여도 단어의 뉘앙스의 차이 때문에 합격이 갈리는 경우도 있다고 들었다. 이에 착안하여 자기소개서의 단어들을 바꿔주는 프로젝트를 설정하게 되었다.

1. **과제 구현**

Python 3.6을 기반으로 하여 Ubuntu환경에서 코드 작업을 진행하였으며 여러 Open Source Library들을 사용하여 과제를 구현하였다. KoNlpy의 mecab라이브러리가 window환경에서는 지원하지 않아 WSL(Windows Subsystem for Linux)나 Ubuntu 18.04 가 설치된 데스크탑 등을 활용하여 과제를 구형하였다.

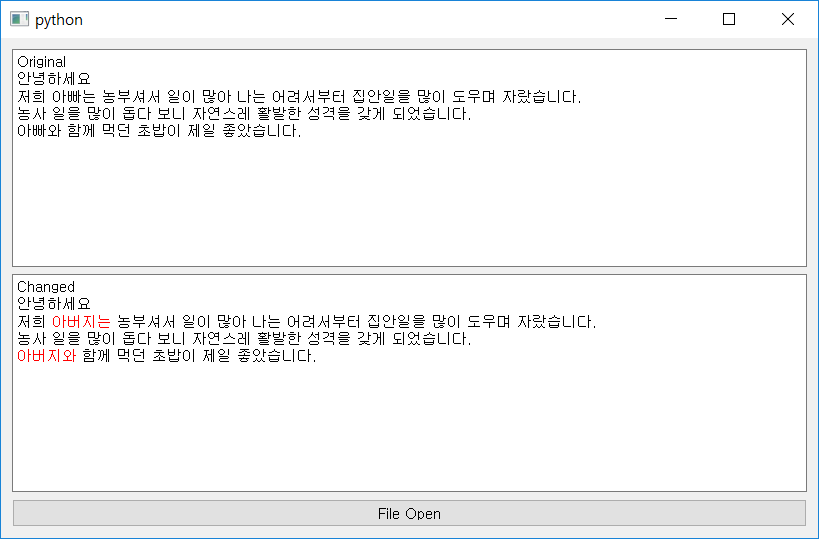
강민수 학우와 김상헌 학우가 형태소 및 어근 분석 부분을 담당하여 KoNlpy 오픈소스를 활용하여 자기소개서의 형태소와 어근을 분리하는 코드인 mecab.py 코드를 작성하고 입력받은 자기소개서의 형태소와 어근을 분리하고 추천 단어 리스트인 recommend.csv에서 해당 단어가 있는지, 있다면 현재의 단어와 유사한 단어의 랭킹을 비교하여 교환해주는 python script를 작성하였다.

임기찬 학우는 유의어 데이터를 찾아서 유의어 사전에서 우리가 필요한 데이터를 사용하기 위해 데이터를 정제하는 작업을 진행하는 extract.py와 ExtractRelatedWords.py python scirpt를 작성하였다.

오준오 학우는 자료의 핵심인 자소서 샘플 데이터를 찾아 Chrome driver와 Ope Source인 셀레니움을 활용하여 자료를 crawling하는 python script인 crawling.py를 작성하였다.

천영재 학우는 오픈소스 라이브러리 중 파이썬 시각화 툴인 PyQt를 사용하여 결과물을 시각화 하였으며 강민수 학우와 김상헌 학우의 코드를 사용하여 바뀐 단어들을 색칠하는 방식으로 시각화 부분을 구현하였다.

1. **결과 화면**



간단한 예시 화면을 보여주기 위해 짧은 글을 작성하여 테스트를 한 결과 화면이다.