

형태소 변환기를 사용한 존댓말 반말 변환기

- **설계의 목적:** 한국어는 교착어로, 높임말과 반말이 존재하며 이는 상황과 장소에 따라 다르게 사용됩니다. 의미는 같으나 다양한 어체로 사용이 가능하며 이로 인해 다른 언어보다 번역이 까다롭습니다. 따라서 저희 팀은 딥러닝 모델에 한국어 데이터를 학습시킬 때, 모든 한국어를 높임말 또는 반말로 통일하여 학습을 시켰을 때 번역의 품질이 향상되는지 확인해보고, 번역된 결과를 높임말 또는 반말로 통일시켜 사용자가 결과 문장을 읽을 때의 가독성을 높이하고자 높임말 반말 변환 코드를 개발하였습니다.

형태소 변환기를 사용한 존댓말 반말 변환기

- **구현 원리:** 한국어 존댓말 반말 변환기를 만드는 것을 구현하는 것에 있어, 존댓말과 반말의 차이를 만드는 것이 문장의 특정 위치 (또는 형태소)에 나타나는 특정 어휘들에 의해 일어난다고 생각했습니다. 그래서 형태소 분석기를 사용하여 문장의 구성 요소들과 이들 각각에 대응하는 형태소를 알아낸 뒤 존댓말, 반말 어휘가 나타나는 ‘어말 어미’ 등에 집중하여 이에 대응하는 어휘들을 교체하는 작업을 진행하였습니다.

높임말 반말 변환 Flow

Input: '우리는 이제 개강을 한다.'



높임말 반말 변환기 모듈

Step1. 띄어쓰기 단위로 문장을 분리한다. 형태소 tag도 띄어쓰기 단위로 분리시킨다.

['우리는', '이제', '개강을', '한다.']
['NP/JX', 'MAG', 'NNG/JKO', 'VV+EF/SF']

Step2. 다시 더 작은 단위인 자모 단위로 분리한 다음, 종결어미를 찾아낸다. (자모 분리는 Api 사용)

['ㅇㅏㄹ | ㄴㅡㄴ', 'ㅇ | ㅈ ㅅ', 'ㄱ ㅅ ㅈ | ㅇㅇㅡㄴ', 'ㅎ | .']
['NP/JX', 'MAG', 'NNG/JKO', 'VV/SF'], ['ㄴ ㄷ | ']

높임말 반말 변환 Flow

높임말 반말 변환기 모듈

Step3. 종결어미가 'ㄴ다'로, '다' 앞에 글자에 종성이 있기에 'ㅂ니다'로 종결어미를 변경한다. 종결어미가 없었다면 '습니다'로 변경될 것이다. 이 step에서 한국어의 문법적인 요소를 고려하여 종결어미를 변경한다.

['ㄴ다'] → ['ㅂ니다']

Step4. 변경된 종결어미를 기존의 문장에 결합한 다음 자모 단위로 나누어진 문장을 합쳐준다.

['우리는 이제 개강을 합니다.']
['NP/JX', 'MAG', 'NNG/JKO', 'VV/SF']

높임말 반말 변환 Flow

높임말 반말 변환기 모듈

- 높임말에서 반말로 변경하는 방법도 유사하나, 용언의 활용도 고려해야 하기에 이에 대한 규칙을 정해 변경하도록 하고 있습니다.
- 한국어의 하십시오체는 해라체로 변경되고, 해요체는 해체로 변경되도록 하였습니다. 하십시오체와 해라체는 같은 격식체이고, 해요체와 해체는 비격식체이기에 같은 어체 사이에서 변경이 이루어지도록 하였습니다.



output: '우리는 이제 개강을 합니다.'