

Lab Meeting

Corpus로 어말 어미 변환

허재무(21700784) 김주환(21700165)

20201.09.14

띄어쓰기 문제 해결

highlow_prototype3.ipynb

- 기존의 변환

```
the changer starts!  
the jamo starts!  
밥을먹었다.
```

- PyKoSpacing을 적용 후

```
the changer starts!  
the jamo starts!  
밥을 먹었다.
```

Mecab 변환기 적용

highlow_prototype3.ipynb

- 기존의 Hannanum의 경우 일부 형태소가 사라져 제대로 된 변환이 안되었습니다.
- Mecab을 적용하여 형태소 분리 후 합칠 때 형태소가 사라지지 않았으며 현재 확인했을 때에는 문제가 없었습니다.

```
txt = '호랑이가 왔습니다.'  
spacing = Spacing()  
ch = Changer()  
tt = ch.processText(txt)  
print(tt)
```

the changer starts!
the jamo starts!
호랑이가 오압니다.

```
txt = '호랑이가 왔습니다.'  
spacing = Spacing()  
ch = Changer()  
tt = ch.processText(txt)  
print(tt)
```

the changer starts!
the jamo starts!
호랑이가 왔다.

Corpus를 바탕으로 Dictionary 추가

highlow_prototype3.ipynb

- ‘주세요’, ‘오세요’같은 경우 변하는 것은 동사입니다. 이때 ‘주’를 무조건 ‘줘’에서 찾는 것이 아닌 다음의 tag EP나 EF가 ‘세’일 경우 그 때 변환을 합니다.