

Lab Meeting

반말 높임말 - 형태소 분석기 적용 후 어미 변환

허재무(21700784), 김주환(21700165)

2021.09.07

기존의 변환기 코드 Review

highlow_prototype1.ipynb

- 각각의 태그 중 어말어미 중 EF(종결어미), EP(선어말 어미)와 일치한다면 그 형태소를 자모로 분리한다.
- 자모단위로 분리된 형태소가 존댓말이 정의된 사전에 존재한다면 그 사전에 나와있는 반말로 변환한다.

```
from konlpy.tag import Hannanum  
han = Hannanum( )
```

기존의 변환기 코드의 문제점

highlow_prototype1.ipynb

- '합니다' -> '하'와 'ㅂ니다'
- 이때 'ㅎ, ㅏ, ㅓ', 'ㅂ, ㅓ, ㅣ, ㅕ, ㅓ'로 자모가 분리되고 반말로 변환하고 합치는 과정에서 'ㅂ'은 종성임에도 불구하고 초성으로 인식되어 제대로 된 변환이 안될 가능성이 있다.

```
-----  
ValueError                                Traceback (most recent call last)  
<ipython-input-3-95cdb7ce9061> in <module>  
    1 txt = '전 좋아합니다'  
    2 ch = Changer()  
----> 3 tt = ch.processText(txt)  
    4 print(tt)  
  
<ipython-input-2-247ed7e33412> in processText(self, stc)  
    189     def processText(self, stc):  
    190         result = stc  
--> 191         res = self.high_low(result)  
    192         return res  
    193  
  
<ipython-input-2-247ed7e33412> in high_low(self, stc)  
    175         res = str2li(res)  
    176         res = to2dim(res)  
--> 177         res = makeone(res)  
    178         lis_word[i] = res  
    179         print(lis_word[i])  
  
<ipython-input-2-247ed7e33412> in makeone(input)  
    88         ind = ord('가')  
    89         ind += CHOSUNG_LIST.index(i[0])*588  
----> 90         ind += JONGSUNG_LIST.index(i[1])*28  
    91         ind += JONGSUNG_LIST.index(i[2])  
    92         result = result+chr(ind)  
  
ValueError: 'ㅕ' is not in list
```

기존의 변환기 코드 보완

highlow_prototype2.ipynb

- 맨 앞에 있는 element가 반드시 초성이어야 한다
 - > hangul_utils에서 제공하는 분리, 결합 함수를 사용하여 이 문제를 해결
- 형태소 유형에 따른 mapping도 구현을 하였다.