

Lab Meeting

띄어쓰기 및 종결어미 이슈 처리

김주환(21700165) 허재무(21700784)

2021.09.28

띄어쓰기 이슈 대안

highlow_prototype6.ipynb

- 원래 문장의 띄어쓰기를 가져오기 위한 방법들을 시도하였습니다.
- 기존 문장의 space가 위치한 index를 기억했다가 반말로 바꾸고 다시 문장의 space 위치에 space(' ')을 집어넣는 방식을 시도하였습니다.

‘- 까’ 종결어미 처리

highlow_prototype6.ipynb

- 원래 문장의 띄어쓰기인 ‘ ‘을 가져오기 위한 방법을 시도하였습니다.
- 기존 문장의 space의 index를 기억했다가 반말로 바꾸고 문장의 space 위치에 space를 집어넣는 방식입니다.

‘- 까’ 종결어미 처리

highlow_prototype6.ipynb

- ‘-까’로 끝나는 종결어미로 '습니까'가 있었습니다. 이는 우선 '-어'로 바꾸도록 사전을 추가하였습니다.
 - 밥 먹었습니까? -> 밥 먹었어?
 - 괜찮습니까? -> 괜찮어?
- 존칭 선어말 어미와 종결어미의 '십니까'의 경우 선어말 어미 ‘시’와 ‘ㅂ니까’를 따로 고려하였습니다. ‘시’를 ‘-니’로 변환하였습니다.
 - 바쁘십니까? 뭐하십니까? -> 바쁘니? 뭐하니?

‘- 까’ 종결어미 처리

highlow_prototype6.ipynb

```
[ '10/SN', '분/NNBC', '정도/NNG', '기다리/VV', '실/EP+ETM', '수/NNB', '있/VV', '습니까/EF', '?/SF' ]  
[ '10/SN', '세/NNBC', '여자/NNG', '아이/NNG', '의/JKG', '선물/NNG', '로/JKB', '는/JX', '무엇/NP', '이/JKS', '좋/VA', '을/ET  
M', '것/NNB', '같/VA', '습니까/EF', '?/SF' ]  
[ '10/SN', '시/NNBC', '정각/NNG', '이/VCP', '에요/EF', './SF' ]  
[ '10/SN', '시/NNBC', '정각/NNG', '이/VCP', '에요/EF', './SF' ]  
[ '10/SN', '시/NNBC', '쯤/XSN', '에/JKB', '목욕/NNG', '해요/XSV+EF', './SF' ]  
[ '10/SN', '일/NNBC', '에/JKB', '는/JX', '바쁘/VA', '습니까/EP+EF', '?/SF' ]
```

10분정도 기다리닐 수 있어?

10세 여자 아이의 선물로는 무엇이 좋을 것 같어?

10시 정각이야.

10시 정각이야.

10시쯤에 목욕해.

10일에는 바쁘니?

‘- 까’ 종결어미 처리

highlow_prototype6.ipynb

```
[ '15/SN', '분/NNBC', '쯤/NNG', '뒤/NNG', '에/JKB', '내/NP+JKB', '방/NNG', '으로/JKB', '와/VV+EC', '주/VX', '실래요/EP+EF',  
'?/SF' ]
```

15분 쯤 뒤에 내 방으로 와 주닐래?
15세기때문에 술은 마실 수 없다.
15일이야.
17일은 몇 요일이야?
1981년에 태어났어.

와 주실래요? -> 와 주 + 니 + 래 ?
이용하십시오 -> 이용하 + 니 + 니 + 오.

‘-까’ 종결어미 처리

개선 방향

- VV(동사)를 계속 추가하는 것은 동사의 수가 많기에 비효율적이라 생각합니다.
- 코퍼스의 데이터를 확인해보면서 제대로 변환이 안되는 부분을 확인했을 때 제한적인 어휘들 안에서만 나타난다면 사전을 추가할 것이고, 아니면 예외처리를 추가할 것입니다.
- 코퍼스의 데이터를 확인해보면서 제대로 변환이 안되는 부분을 확인했을 때 제한적인 어휘들 안에서만 나타난다면 사전을 추가할 것이고, 아니면 예외처리를 추가할 것입니다.