

Lab Meeting

사전 정교화 및 마침표 이슈 해결

허재무(21700784) 김준태(21600150) 김정희(21700156) 김주환(21700165)
2021.11.02

다중 문장 처리- 반말->높임말

highlow_prototype9.ipynb

- 한 데이터에서 문장이 2개 이상 존재하면 뒤에 문장만 처리했습니다.
- findTargetMorphology 함수를 추가해서 특정한 tag만 찾아서 바꾸도록 하였습니다.
- 예를 들어, 종결어미인 (EF)를 데이터에서 찾아서 바꿉니다.

사전 정교화 - 반말->높임말

highlow_prototype9.ipynb

- 사전을 연속적으로 참고하는 경우가 생겨서 이를 사전을 한번만 참고하도록 변경하였습니다.

싫다고.
싫습니까고요.

싫다고. 몇 번을 말해?
싫다고요. 몇 번을 말해요?

```
##added by JM
flag = 0
for i in range(len(dic)):
    if dic[i][0] in res:
        flag = 1
    if len(dic[i])>2:
        res = re.sub(dic[i][0],dic[i][1],res)
    else:
        res = re.sub(dic[i][0],dic[i][1],res)
    if flag==1:
        break
jam.jamo[ind] = res
```

앞으로의 방향성

highlow_prototype9.ipynb

- 의문문인지 평서문인지에 따라 변환되는 형태소를 다르게 처리하거나 아니면 ‘-요’로 통일을 해야 합니다.
- 되? / 돼?로 끝나는 문장 처리
- ‘~게.’, ‘~니’ 로 끝나는 반말 처리 -> 종결어미로 끝나지 않을 가능성이 있습니다.
- 종결어미로 끝나지 않는 문장의 경우를 처리해주어야 합니다.
 - ‘뭐라든?’, ‘빨리 하라는데.’ 등등의 문장이 그 예시입니다.

CD는 어디에 있어요.

DVD는 화질이 좋아서 좋습니다?

DVD를 사는 것은 영원히 소장할 수 있기 때문입니다.

DVD플레이어를 갖고 싶어요.

E-메일을 보내줘서 고마워요.

EPA 는 나쁜 콜레스테롤을 줄여줍니다.

E메일의 주고받기와 인터넷을 할 줄 몰라요.