

Lab Meeting

높임말-> 반말 데이터 파일 생성

김주환(21700165) 허재무(21700784)
2021.10.12

기존 높임말 -> 반말 보완

highlow_prototype8.ipynb

- 지난 주의 문제
 - ‘바꿔주세요’ => ‘바꿔줘나’로 변경되었습니다.
 - 이는 ‘세요’를 ‘나’로 바꿔도록 사전이 구성되어 있었기 때문입니다.
 - ‘세요’의 경우 의문문인 경우와 평서문의 경우 반말의 형태를 다르게 해주어야 합니다.

기존 높임말 -> 반말 보완

highlow_prototype8.ipynb

- 예외 사전에 들어가기 전 예외 인덱스를 찾는 방법은 너무 복잡하였습니다.
- 이를 해결하기 위해 규칙을 정해서 찾으도록 하였습니다. 변형이 많은 경우 'special'로 처리합니다.

```
for j in range(len(dic)):
    if self.isExcept(dic[j])==1:
        #ind = self.indicator(i,jam.jamo,lis_tag,EXC_word, EXC_tags)
        for q in range(len(EXC_deal)):
            if dic[j][0]==EXC_deal[q][0]:
                for r in range(1,len(EXC_deal[q])):
                    if jam.jamo[i-1][-1]==EXC_deal[q][r][0]:
                        res = re.sub(dic[j][0],EXC_deal[q][r][1],res);
        #res = re.sub(dic[j][0],dic[j][ind],res)
```

기존 높임말 -> 반말 보완

highlow_prototype8.ipynb

- 예외로 처리를 해 문제점을 보완하였습니다.

```
EP_EF = [  
    [['ㄱ', '세', 'ㅇ', 'ㅍ'], ['special']]  
    #-시오 Tammy  
    , [['ㅡ', 'ㅅ', 'ㅣ', 'ㅁ', 'ㅅ', 'ㅣ', 'ㅇ', 'ㅊ'], ['ㅈ']]  
    , [['ㅅ', 'ㅣ', 'ㅁ', 'ㅅ', 'ㅣ', 'ㅇ', 'ㅊ'], ['ㄹ', 'ㅈ']],  
    [['ㅅ', 'ㅣ', 'ㅁ', 'ㄴ', 'ㅣ', 'ㅁ', 'ㅈ'], ['ㄴ', 'ㅣ']],  
    [['ㅅ', 'ㅣ', 'ㄹ', 'ㄹ', 'ㅅ', 'ㅇ', 'ㅍ'], ['ㄹ', 'ㄹ', 'ㅅ']],  
]
```

종결어미로 '세요'가 오면 special로 처리합니다.

```
EXC_deal = [  
    ['ㄱ세ㅇㅍ', ['ㅈ', 'ㅣ'], ['ㅈ', 'ㅈ'], ['ㅡ', 'ㅈ']]  
]
```

어서 하세요 -> 어서 하 + ㅣ => 어서해
빌려주세요 -> 빌려주 + ㅈ => 빌려줘
밥 먹으세요 -> 밥 먹ㅇ + ㅈ => 밥 먹어

기존 높임말 -> 반말 데이터 파일 생성

hgu_clean.low.kr.shuf

- 높임말을 반말로 변환 후 이를 파일에 저장하였습니다.

26일이야?
27일은 괜찮나?
중단시켜서 미안해.
중단시켜서 미안해.
아, 이런! 도저히 생각이 안 나는걸.
Joe's Café 라는 레스토랑에 가고 싶은데, 알고 있어?
보다 나은 건강과 생활 회사의 글렌 토마스이다.
스타워즈는 어때?
데이터 코디네이터나 데이터 매니저가 어떨까?
화성에서 온 로봇이! 좋아! 특수 효과가 대단하대.
팩스를 보내줘서 감사한다.
11월30일 날짜로 팩스를 보내줘서 감사한다.
자격을 취득하기 위해 공부 중이야.
간호사 자격을 취득하기 위해 공부 중이야.
국가시험을 향해 공부 중이야.
공무원에 되기 위한 국가시험을 향해 공부 중이야.
잘 됐어. 그럼 이제 집에 가도 되나?
미안한다. 그녀는 안에 없다. 메모 남겨 드릴까?
내려 줄래.
올려 줄래.
작은 목소리.
스미스 선생님. 저 화장실에 가야겠는데.
화장실이 정말 급해!
그럼 총수입의 얼마나 되는 거야?
0136521798이야.
1,199달러야?
100그램에 비타민C의 함유량은 레몬보다 높다.
100미터 더 가서 좌회전해.
105번이 그 근처에 서는 것 같은데, 확실하진 않아.
넨시에게 물어 보세요. 넨시는 나보다 더 자주 그 버스를 이용하거든.
