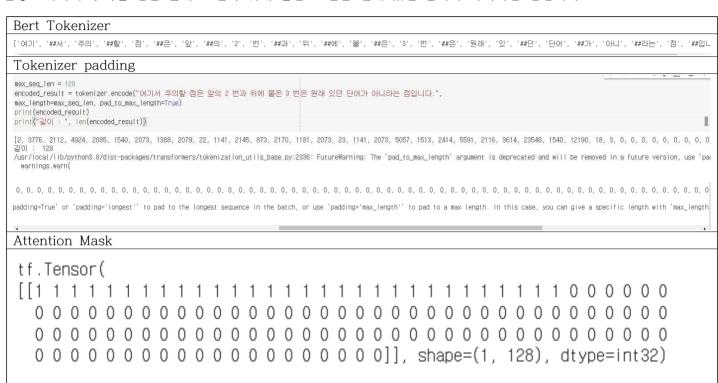
과제 #9

201904479 컴퓨터 전자 시스템 학과 임서연

1. 이 자료의 8~12쪽을 참고하여 다음 문장의 토큰수를 최대 128로 하고 tokenize 결과 (9쪽 형식), 인코딩 결과(11쪽 형식), attention mask(12쪽 형식)를 구하라. 문장: '여기서 주의할 점은 앞의 2 번과 뒤에 붙은 3 번은 원래 있던 단어가 아니라는 점입니다.'



2. [마스크 실습] 이 자료의 15~16쪽을 참고하여 다음 문장에 대한 마스크 처리 결과를 구하라. 문장: '이 영화의 초점은 [MASK]다.'

```
inputs = tokenizer('이 영화의 초점은 [MASK]다.', return_tensors='tf')
print(inputs['input ids'])
[{'score': 0.05608559027314186,
  'token': 3657,
  'token_str': '하나',
  'sequence': '이 영화의 초점은 하나 다.'},
 {'score': 0.04044365510344505.
  'token': 1504,
  'token_str': '0|',
  'sequence': '이 영화의 초점은 이 다.'},
 {'score': 0.03929238021373749.
  'token': 1417.
  'token str': '었'.
  'sequence': '이 영화의 초점은 었 다.'},
 {'score': 0.03794778138399124.
  'token': 5969.
  'token_str': '남녀',
  'sequence': '이 영화의 초점은 남녀 다.'},
 {'score': 0.029111366719007492.
  'token': 35,
  'token_str': '?',
  'sequence': '이 영화의 초점은? 다.'}]
```

3. [네이버 영화평] 이 자료의 17~25쪽의 프로그램들을 순차적으로 실행하여 영화평 분류기를 수행한다. 계산시간이 많이 걸리므로 훈련용, 테스트용 영화평을 각각 15000개와 5000개로 줄여서 실행한다. 훈련을 마친 후 25쪽과 유사하게 다음 영화평에 대한 결과를 구하라. 영화평: '시간은 그럭저럭 때울 수 있기도', '감독은 노력했지만 결과는 그다지...'