

메모1: 문자열을 읽어와서 어절 단위로 쪼개기

작성자: 이선환

작업 시작일: 2021년 6월 25일

텍스트 파일로 문자열을 입력받고, 띄어쓰기나 구두점(마침표, 물음표, 느낌표 등)으로 단위를 구분하기.
도스(Dos)의 프롬프트에서 파이썬을 실행한 결과를 정리하면 아래와 같다.

```
C:\Wpython\WDay5\WProject-of-lsh>python
Python 3.9.5 (tags/v3.9.5:0a7dcdb, May 3 2021, 17:27:52) [MSC v.1928 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> f = open('source/2.txt', 'r')
>>> r = f.read()
>>> r
'lo, world!!\nI am lee seon hwan!!\nnice to meet you!!\Wntoo!!!!'
>>> r.split(' ')
['lo,', 'world!!\nI', 'am', 'lee', 'seon', 'hwan!!\nnice', 'to', 'meet', 'you!!\Wntoo!!!!']
>>> r.split(' ' or '\n')
>>> r.split('\n')
['lo, world!!', 'I am lee seon hwan!!', 'nice to meet you!!', 'too!!!!']
>>> a = r.split('\n')
>>> a
['lo, world!!', 'I am lee seon hwan!!', 'nice to meet you!!', 'too!!!!']
>>> a.split(' ')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'list' object has no attribute 'split'
>>> a[0]
'lo, world!!'
>>> a[0].split(' ')
['lo,', 'world!!']
```

결론: 일단, 일정한 범위나 단위로 문자열을 가져온 후, split()으로 다시 쪼개어서 어절 단위로 만들어야 한다. 그렇게 어절 단위로 나눈 다음 이를 기준으로 딕셔너리(dictionary)의 키(key) 값으로 다시 넣어 주어야 되겠다.

작업 마무리일: 2021년 6월 25일

참고 자료: