

# 실습: 파이썬 전체 조망

비교적 작은 규모의 실제 예제로 파이썬 프로그래밍 전체를 조망

1. 주석
2. import 모듈
3. requests
4. split()
5. 파이썬 딕셔너리(dictionary)
6. Counter, Counter.most\_common()

## word counter

```
# word counter
import requests
from collections import Counter

r = requests.get('https://www.gnu.org/licenses/gpl.txt')
words = r.text.lower().split()
word_dict = {}

for key in words:
    word_dict[key]=word_dict.get(key,0)+1

count = Counter(word_dict)
common = count.most_common(20)
print (common)
```

```
[('the', 344), ('of', 219), ('to', 188), ('a', 178), ('or', 142), ('you', 123), ('and', 91), ('that', 89), ('for', 83), ('this', 83), ('in', 75), ('is', 69), ('license', 63), ('work', 60), ('not', 49), ('any', 48), ('if', 45), ('under', 44), ('by', 43), ('with', 43)]
```

## 주석

```
# comment
```

## 모듈 가져오기

```
import requests
from collections import Counter
```

## requests 모듈

requests 특정 URL로 부터 페이지를 가져온다. get() 혹은 post()를 주로 사용한다. import 혹은 from ~ import 를 사용한다.

```
r = requests.get('https://www.gnu.org/licenses/gpl.txt')
print (r.text[:400])
```

GNU GENERAL PUBLIC LICENSE  
Version 3, 29 June 2007

Copyright (C) 2007 Free Software Foundation, Inc. <<http://fsf.org/>>  
Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

### Preamble

The GNU General Public License is a free, copyleft license for  
software an

## str.lower(), str.split()

lower() 함수는 문자열을 소문자로 바꾼다. split()은 공백문자(whitespace)를 기준으로 분리하여 리스트로 반환

```
words = r.text.lower().split()
print(len(words))
words[:10]
```

5644

```
['gnu',
 'general',
 'public',
 'license',
 'version',
 '3,',
 '29',
 'june',
 '2007',
 'copyright']
```

## 딕셔너리

- 딕셔너리를 사용해 단어 수 세어 합산 하기
- for i in list

```
word_dict = {}
```

```
for key in words:
    word_dict[key] = word_dict.get(key,0)+1
```

## 개수 세기, TOP-N 구하기

- Counter() 딕셔너리에 저장된 값을 카운트
- Counter.most\_common(20) 최상위 20개

```
count = Counter(word_dict)
common = count.most_common(10)
common
```

```
[('the', 344),
 ('of', 219),
 ('to', 188),
 ('a', 178),
 ('or', 142),
 ('you', 123),
 ('and', 91),
 ('that', 89),
 ('for', 83),
 ('this', 83)]
```