



# 자연어 처리 실습 1

과목명 자연어 처리

담당교수 김낙현교수님

제출일 20210910

전공 컴퓨터전자시스템

학번 201904458

이름 이준용



**한국외국어대학교**  
HANKUK UNIVERSITY OF FOREIGN STUDIES

```
!pip install konlpy
```

```
Requirement already satisfied: konlpy in /usr/local/lib/python3.7/dist-packages (0.5.2)
Requirement already satisfied: beautifulsoup4==4.6.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: colorama in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: JPype1>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: lxml>=4.1.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: tweepy>=3.7.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: numpy>=1.6 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: requests[socks]>=2.11.1 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from konlpy)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.7/dist-packages (from konlpy)
```

```
# 메밀꽃.txt에서 나온 횟수가 5번이상인 단어 출력
```

```
from konlpy.tag import Kkma
```

```
from collections import Counter
```

```
infile = open("메밀꽃.txt", encoding='utf-8')
```

```
data=infile.read()
```

```
kkma = Kkma()
```

```
words=kkma.morphs(data)
```

```
vocab=Counter(words)
```

```
print(vocab)
```

```
vo=[print(f'-vlaue {k} appears multiple times ({v} times)') for k,v in  
vocab.items() if v > 4]
```

```
del vo[:1]
```

```
-vlaue 그 appears multiple times (18 times)
```

```
-vlaue 으면 appears multiple times (5 times)
```

```
-vlaue 같이 appears multiple times (10 times)
```

```
-vlaue 계집 appears multiple times (12 times)
```

```
-vlaue 로 appears multiple times (15 times)
```

```
-vlaue ..... appears multiple times (10 times)
```

```
-vlaue 집 appears multiple times (7 times)
```

```
-vlaue 말 appears multiple times (12 times)
```

```
-vlaue 야 appears multiple times (16 times)
```

```
-vlaue 뜻이 appears multiple times (5 times)
```

```
-vlaue 구 appears multiple times (7 times)
```

```
-vlaue 그럴 appears multiple times (9 times)
```

```
-vlaue 동이 appears multiple times (19 times)
```

```
-vlaue 세 appears multiple times (7 times)
```

```
-vlaue 충춇집 appears multiple times (5 times)
```

```
-vlaue 녀석 appears multiple times (9 times)
```

```
-vlaue 알 appears multiple times (5 times)
```

```
-vlaue 길 appears multiple times (6 times)
```

```
-vlaue 만 appears multiple times (14 times)
-vlaue 수 appears multiple times (20 times)
-vlaue 마음 appears multiple times (5 times)
-vlaue 버리 appears multiple times (6 times)
-vlaue 때 appears multiple times (16 times)
-vlaue 치 appears multiple times (6 times)
-vlaue 부터 appears multiple times (7 times)
-vlaue 앞 appears multiple times (5 times)
-vlaue 면 appears multiple times (12 times)
-vlaue 하나 appears multiple times (6 times)
-vlaue 겠 appears multiple times (11 times)
-vlaue 란 appears multiple times (5 times)
-vlaue 니 appears multiple times (10 times)
-vlaue 같 appears multiple times (8 times)
-vlaue 면서 appears multiple times (5 times)
-vlaue 젊 appears multiple times (6 times)
-vlaue 그것 appears multiple times (5 times)
-vlaue 러 appears multiple times (6 times)
-vlaue 오 appears multiple times (5 times)
-vlaue 짐승 appears multiple times (10 times)
-vlaue 눈 appears multiple times (6 times)
-vlaue 나귀 appears multiple times (15 times)
-vlaue 젓 appears multiple times (7 times)
-vlaue 까 appears multiple times (5 times)
-vlaue 울 appears multiple times (6 times)
-vlaue 속 appears multiple times (7 times)
-vlaue 아이 appears multiple times (7 times)
-vlaue 아니 appears multiple times (5 times)
-vlaue 저 appears multiple times (5 times)
-vlaue 제천 appears multiple times (7 times)
-vlaue 라고 appears multiple times (6 times)
-vlaue 물 appears multiple times (13 times)
-vlaue 다고 appears multiple times (6 times)
-vlaue 평 appears multiple times (6 times)
-vlaue 지금 appears multiple times (7 times)
-vlaue 이야기 appears multiple times (7 times)
-vlaue 처녀 appears multiple times (6 times)
-vlaue 어머니 appears multiple times (5 times)
-vlaue 쯤 appears multiple times (6 times)
```

```
# 소나기.txt에서 나온 횟수가 5번이상인 단어 출력
from konlpy.tag import Kkma
from collections import Counter
infile = open("소나기.txt", encoding='utf-8')
data=infile.read()

kkma = Kkma()
words=kkma.morphs(data)

vocab=Counter(words)
print(vocab)
vo=[print(f'-vlaue {k} appears multiple times ({v} times)') for k,v in
    vocab.items() if v > 4]
del vo[:1]
```

```

-vlaue 알 appears multiple times (22 times)
-vlaue 쪽 appears multiple times (5 times)
-vlaue 안 appears multiple times (8 times)
-vlaue 의 appears multiple times (35 times)
-vlaue 같 appears multiple times (16 times)
-vlaue 뵈 appears multiple times (7 times)
-vlaue 덜 appears multiple times (5 times)
-vlaue 집 appears multiple times (7 times)
-vlaue 주머니 appears multiple times (5 times)
-vlaue 일 appears multiple times (6 times)
-vlaue □ appears multiple times (5 times)
-vlaue 잡 appears multiple times (6 times)
-vlaue 전 appears multiple times (9 times)
-vlaue 더 appears multiple times (17 times)
-vlaue 두 appears multiple times (6 times)
-vlaue 으로 appears multiple times (27 times)
-vlaue 번 appears multiple times (5 times)
-vlaue 말 appears multiple times (12 times)
-vlaue 느냐 appears multiple times (5 times)
-vlaue ' appears multiple times (12 times)
-vlaue 어서 appears multiple times (9 times)
-vlaue 없 appears multiple times (8 times)
-vlaue 없이 appears multiple times (5 times)
-vlaue 소리 appears multiple times (6 times)
-vlaue 오 appears multiple times (12 times)
-vlaue 니 appears multiple times (9 times)
-vlaue 보이 appears multiple times (5 times)
-vlaue 못 appears multiple times (6 times)
-vlaue 뜰 appears multiple times (10 times)
-vlaue 이것 appears multiple times (8 times)
-vlaue ? appears multiple times (16 times)
-vlaue 자기 appears multiple times (6 times)
-vlaue 눈 appears multiple times (6 times)
-vlaue 참 appears multiple times (8 times)
-vlaue 산 appears multiple times (6 times)

-vlaue 우리 appears multiple times (5 times)
-vlaue 멀 appears multiple times (5 times)
-vlaue 면 appears multiple times (5 times)
-vlaue 허수아비 appears multiple times (6 times)
-vlaue 들 appears multiple times (7 times)
-vlaue 대 appears multiple times (5 times)
-vlaue 달려가 appears multiple times (5 times)
-vlaue 그 appears multiple times (14 times)
-vlaue ㄴ다는 appears multiple times (5 times)
-vlaue 곳 appears multiple times (6 times)
-vlaue 밑 appears multiple times (6 times)
-vlaue 수숫단 appears multiple times (5 times)
-vlaue 꽃 appears multiple times (14 times)
-vlaue …… appears multiple times (13 times)
-vlaue 그럴 appears multiple times (5 times)
-vlaue 크 appears multiple times (5 times)
-vlaue 가지 appears multiple times (5 times)
-vlaue 송아지 appears multiple times (6 times)
-vlaue 비 appears multiple times (7 times)
-vlaue 이사 appears multiple times (6 times)
-vlaue 호두 appears multiple times (6 times)
-vlaue 아버지 appears multiple times (5 times)

```

```
# 메밀꽃.TXT에서 줄풍어를 제거하고 나온 윗수가 5번이상인 단어 출력
from konlpy.tag import Kkma
from collections import Counter
infile = open("메밀꽃.txt", encoding='utf-8')
data=infile.read()
stopfile = open("불용어.txt", encoding='utf-8')
stopwords = stopfile.read()

kkma = Kkma()
words=kkma.morphs(data)
stopwords=kkma.morphs(stopwords)

result=[]
for w in words:
    if w not in stopwords:
        result.append(w)

vocab=Counter(result)
print(vocab)
vo=[print(f'-vlaue {k} apppears multiple times ({v} times)') for k,v in
    vocab.items() if v > 4]
del vo[:1]
```

```
Counter({'으나': 25, '생원': 25, '동이': 19, '나귀': 15, '장': 13, '꿀': 13, '이나': 12, '계
-vlaue 장 apppears multiple times (13 times)
-vlaue 이란 apppears multiple times (5 times)
-vlaue 장판 apppears multiple times (6 times)
-vlaue 벌써 apppears multiple times (7 times)
-vlaue 뒤 apppears multiple times (8 times)
-vlaue 으나 apppears multiple times (25 times)
-vlaue 이나 apppears multiple times (12 times)
-vlaue 축 apppears multiple times (5 times)
-vlaue ㄴ지 apppears multiple times (6 times)
-vlaue 스럽 apppears multiple times (8 times)
-vlaue 군 apppears multiple times (5 times)
-vlaue 생원 apppears multiple times (25 times)
-vlaue 조선 apppears multiple times (9 times)
-vlaue 달 apppears multiple times (10 times)
-vlaue 봉 apppears multiple times (9 times)
-vlaue 을까 apppears multiple times (5 times)
-vlaue 대화 apppears multiple times (5 times)
-vlaue 밤 apppears multiple times (10 times)
-vlaue 걸 apppears multiple times (8 times)
-vlaue 달이 apppears multiple times (5 times)
-vlaue 소리 apppears multiple times (9 times)
-vlaue 시작 apppears multiple times (10 times)
-vlaue 꿀 apppears multiple times (13 times)
-vlaue 계집 apppears multiple times (12 times)
-vlaue 듯이 apppears multiple times (5 times)
-vlaue 동이 apppears multiple times (19 times)
-vlaue 세 apppears multiple times (7 times)
-vlaue 충춇집 apppears multiple times (5 times)
-vlaue 녀석 apppears multiple times (9 times)
-vlaue 길 apppears multiple times (6 times)
-vlaue 버리 apppears multiple times (6 times)
-vlaue 치 apppears multiple times (6 times)
```

```
-vlaue 란 appears multiple times (5 times)
-vlaue 젊 appears multiple times (6 times)
-vlaue 짐승 appears multiple times (10 times)
-vlaue 눈 appears multiple times (6 times)
-vlaue 나귀 appears multiple times (15 times)
-vlaue 젖 appears multiple times (7 times)
-vlaue 울 appears multiple times (6 times)
-vlaue 재천 appears multiple times (7 times)
-vlaue 라고 appears multiple times (6 times)
-vlaue 다고 appears multiple times (6 times)
-vlaue 평 appears multiple times (6 times)
-vlaue 이야기 appears multiple times (7 times)
-vlaue 처녀 appears multiple times (6 times)
-vlaue 어머니 appears multiple times (5 times)
-vlaue 쪼 appears multiple times (6 times)
```

```
# 소나기.txt에서 불용어를 제거하고 나온 횟수가 5번 이상인 단어 출력
from konlpy.tag import Kkma
from collections import Counter
infile = open("소나기.txt", encoding='utf-8')
data=infile.read()
stopfile = open("불용어.txt", encoding='utf-8')
stopwords = stopfile.read()

kkma = Kkma()
words=kkma.morphs(data)
stopwords=kkma.morphs(stopwords)

result=[]
for w in words:
    if w not in stopwords:
        result.append(w)

vocab=Counter(result)
print(vocab)
vo=[print(f'-vlaue {k} appears multiple times ({v} times)') for k,v in
    vocab.items() if v > 4]
del vo[:1]
```

```
Counter({'소녀': 72, '소년': 49, '날': 14, '꽃': 14, '손': 11, '앗': 10, '들': 10, '개울가':
-vlaue 소년 appears multiple times (49 times)
-vlaue 개울가 appears multiple times (8 times)
-vlaue 소녀 appears multiple times (72 times)
-vlaue 운 appears multiple times (5 times)
-vlaue 초시 appears multiple times (5 times)
-vlaue 개울 appears multiple times (6 times)
-vlaue 손 appears multiple times (11 times)
-vlaue 듯이 appears multiple times (6 times)
-vlaue 돌아오 appears multiple times (7 times)
-vlaue 길 appears multiple times (8 times)
-vlaue 더니 appears multiple times (6 times)
-vlaue 오늘 appears multiple times (5 times)
-vlaue 징검다리 appears multiple times (6 times)
-vlaue 앓 appears multiple times (10 times)
```

-vlaue 버리 appears multiple times (5 times)  
-vlaue 날 appears multiple times (14 times)  
-vlaue 얼굴 appears multiple times (8 times)  
-vlaue 움키 appears multiple times (5 times)  
-vlaue 는지 appears multiple times (6 times)  
-vlaue 건너 appears multiple times (5 times)  
-vlaue 다가 appears multiple times (6 times)  
-vlaue 조약돌 appears multiple times (5 times)  
-vlaue 고는 appears multiple times (8 times)  
-vlaue 바보 appears multiple times (7 times)  
-vlaue 머리 appears multiple times (6 times)  
-vlaue 갈발 appears multiple times (5 times)  
-vlaue 가을 appears multiple times (5 times)  
-vlaue 갈꽃 appears multiple times (7 times)  
-vlaue 다고 appears multiple times (5 times)  
-vlaue 뵈 appears multiple times (7 times)  
-vlaue 덜 appears multiple times (5 times)  
-vlaue 주머니 appears multiple times (5 times)  
-vlaue 잡 appears multiple times (6 times)  
-vlaue 번 appears multiple times (5 times)  
-vlaue 느냐 appears multiple times (5 times)  
-vlaue 없이 appears multiple times (5 times)  
-vlaue 소리 appears multiple times (6 times)  
-vlaue 못 appears multiple times (6 times)  
-vlaue 뜰 appears multiple times (10 times)  
-vlaue 눈 appears multiple times (6 times)  
-vlaue 산 appears multiple times (6 times)  
-vlaue 멀 appears multiple times (5 times)  
-vlaue 허수아비 appears multiple times (6 times)  
-vlaue 달려가 appears multiple times (5 times)  
-vlaue ㄴ다는 appears multiple times (5 times)  
-vlaue 밑 appears multiple times (6 times)  
-vlaue 수숫단 appears multiple times (5 times)  
-vlaue 꽃 appears multiple times (14 times)  
-vlaue 송아지 appears multiple times (6 times)  
-vlaue 이사 appears multiple times (6 times)  
-vlaue 호두 appears multiple times (6 times)  
-vlaue 아버지 appears multiple times (5 times)

✓ 4초 오전 4:48에 완료됨





이쿠 아이고 어 나 우리 저희 따라 의해 올 를 에 의 가 으로 로 에게 뿐이다 의거하여 근거하여 입각하여 기준으로 예하면 예를 들면 예를 들자면 저 소인 소생 저희 지말고 하지마 하지마라 다른 물론 또한 그  
다 뿐만 아니라 만이 아니다 만은 아니다 막론하고 관계없이 그치지 않다 그러나 그런데 하지만 든간에 논하지 않다 따지지 않다 설사 비록 더라도 아니면 만 못하다 하는 편이 낫다 불문하고 향하여 향해서  
타다 오르다 제외하고 이 외에 이 밖에 하여야 비로소 한다면 몰라도 외에도 이곳 여기 부터 기점으로 따라서 할 생각이다 하려고하다 이리하여 그리하여 그렇게 함으로써 하지만 일때 할때 앞에서 중에서  
야한다 일것이다 반드시 할줄알다 할수있다 할수있어 임에 틀림없다 한다면 등 등등 제 겨우 단지 다만 할뿐 덩동 덩그 대해서 대하여 대하면 훨씬 얼마나 얼마만큼 얼마큼 남짓 여 얼마간 약간 다소 좀 조금  
그러나 그렇지만 하지만 이외에도 대해 말하자면 뿐이다 다음에 반대로 반대로 말하자면 이와 반대로 바꾸어서 말하면 바꾸어서 한다면 만약 그렇지않으면 가악 록 딱 빼격거리다 보드득 비격거리다 과당  
여러분 각종 각자 제각기 하도록하다 와 과 그러므로 그래서 고로 한 까닭에 하기 때문에 거니와 이지만 대하여 관하여 관한 과연 실로 아니나다를가 생각한대로 진짜로 한적이있다 하곤하였다 하 하하 허허  
때문에 어찌 하겠는가 무슨 어디 어느곳 더군다나 하물며 더욱이는 어느때 언제 야 이봐 어이 여보시오 흐흐 흥 휴 헉헉 헐떡헐떡 영차 여차 어기여차 공공 아야 앓 아야 쿵쾅 졸졸 짹짹 똑똑 주룩주룩 쇠  
말하면 바꾸어말하자면 혹은 혹시 답다 및 그에 따르는 때가 되어 즉 지든지 설령 가령 하더라도 할지라도 일지라도 지든지 몇 거의 하마터면 인젠 이젠 된바에야 된이상 만큼 어찌됐든 그위에 게다가 점에  
게될것이다 일것이다 비교적 좀 보다더 비하면 시키다 하게하다 할만하다 의해서 연이서 이어서 잇따라 뒤따라 뒤이어 결국 의지하여 기대여 통하여 자마자 더욱더 불구하고 얼마든지 마음대로 주저하지 않  
밖에 안된다 하면된다 그래 그렇지 요컨대 다시 말하자면 바꿔 말하면 즉 구체적으로 말하자면 시작하여 시초에 이상 허 헉 허걱 바와같이 해도좋다 해도된다 게다가 더구나 하물며 와르르 팍 팍 펄렁 동안  
로부터 까지 예하면 했어요 해요 함께 같이 더불어 마저 마저도 양자 모두 습니다 가까스로 하려고하다 즈음하여 다른 다른 방면으로 해봐요 습니까 했어요 말할것도 없고 무릎쓰고 개의치않고 하는것만 못  
들 모 어느것 어느 로써 갖고말하자면 어디 어느쪽 어느것 어느해 어느 년도 라 해도 언젠가 어떤것 어느것 저기 저쪽 저것 그때 그럼 그러면 요만한걸 그래 그때 저것만큼 그저 이르기까지 할 줄 안다 할 힘  
가 차라리 할지언정 할지라도 할망정 할지언정 구토하다 게우다 토하다 메쓰겁다 옆사람 튜 쳇 의거하여 근거하여 의해 따라 힘입어 그 다음 버금 두번째로 기타 첫번째로 나머지는 그중에서 견지에서 형식으  
의해되다 하도록시키다 뿐만아니라 반대로 전후 전자 앞의것 잠시 잠깐 하면서 그렇지만 다음에 그러한즉 그런즉 남들 아무거나 어찌하든지 같다 비슷하다 예컨대 이럴정도로 어떻게 만약 만일 위에서 서  
않는다면 만약에 무엇 무슨 어느 어떤 아래윗 조차 한데 그럼에도 불구하고 여전히 심지어 까지도 조차도 하지 않도록 않기 위하여 때 시각 무렵 시간 동안 어때 어떠한 하여금 네 예 우선 누구 누가 알겠는가  
다 하는 김에 검사검사 하는바 그런 까닭에 한 이유는 그러니 그러니까 때문에 그 너희 그들 너희들 타인 것 것들 너 위하여 공동으로 동시에 하기 위하여 어찌하여 무엇때문에 봉봉 뽁뽁 나 우리 영영 휘익  
하다 하기보다는 차라리 하는 편이 낫다 흐흐 놀라다 상대적으로 말하자면 마치 아니라면 썬 그렇지 않으면 그렇지 않다면 만 그러면 아니었다면 하든지 아니면 이라면 좋아 알았어 하는것도 그만이다 어쩔수  
한편으로는 오자마자 이렇게되면 이와같다면 전부 한마디 한항목 근거로 하기에 아울러 하지 않도록 않기 위해서 이르기까지 이 되다 로 인하여 까닭으로 이유만으로 이로 인하여 그래서 이 때문에 그러므  
을 낼 수 있다 으로 인하여 있다 어떤것 관계가 있다 관련이 있다 연관되다 어떤것들 에 대해 이리하여 그리하여 여부 하기보다는 하느니 하면 할수록 운운 이러이러하다 하구나 하도다 다시말하면 다음으로  
일을 오히려 하기는한데 어떻게 어떻게 어찌됐어 어때 어째서 본대로 자 이 이쪽 여기 이것 이번 이렇게말하자면 이런 이러한 이와 같은 요만큼 요만한 것 얼마 안 되는 것 이만큼 이 정도의 이렇게 많은 것 이  
이 끼익 빼격 따위 와 같은 사람들 부류의 사람들 왜냐하면 중의하나 오직 오로지 에 한하다 하기만 하면 도착하다 까지 미치다 도달하다 정도에 이르다 할 지경이다 결과에 이르다 관해서는 여러분 하고 있  
우에 종합한것과같이 총적으로 보면 총적으로 말하면 총적으로 대로 하다 으로서 참 그만이다 할 따름이다 쿵 탕탕 쿵쿵 뽁 뽁라 아이야 아니와야 응 아이 참나 년 월 일 령 영 일 이 삼 사 오 육 륙 칠  
천구 하나 둘 셋 넷 다섯 여섯 일곱 여덟 아홉 령 영 이 있 하 것 들 그 되 수 이 보 앓 없 나 사람 주 아니 등 갈 우리 때 년 가 한 지 대하 오 말 일 그렇 위하 때문 그것 두 말하 알 그러나 받 못하 일 그런 또  
따르 중 나오 가지 씨 시키 만들 지금 생각하 그러 속 하나 집 살 모르 적 월 데 자신 안 어떤 내 내 경우 명 생각 시간 그녀 다시 이런 앞 보이 번 나 다른 어떻 여자 개 전 들 사실 이렇 점 싶 말 정도 좀 원 잘