입시용 유사 문제 찾기 프로젝트

3조

목차

프로젝트 수행 상황

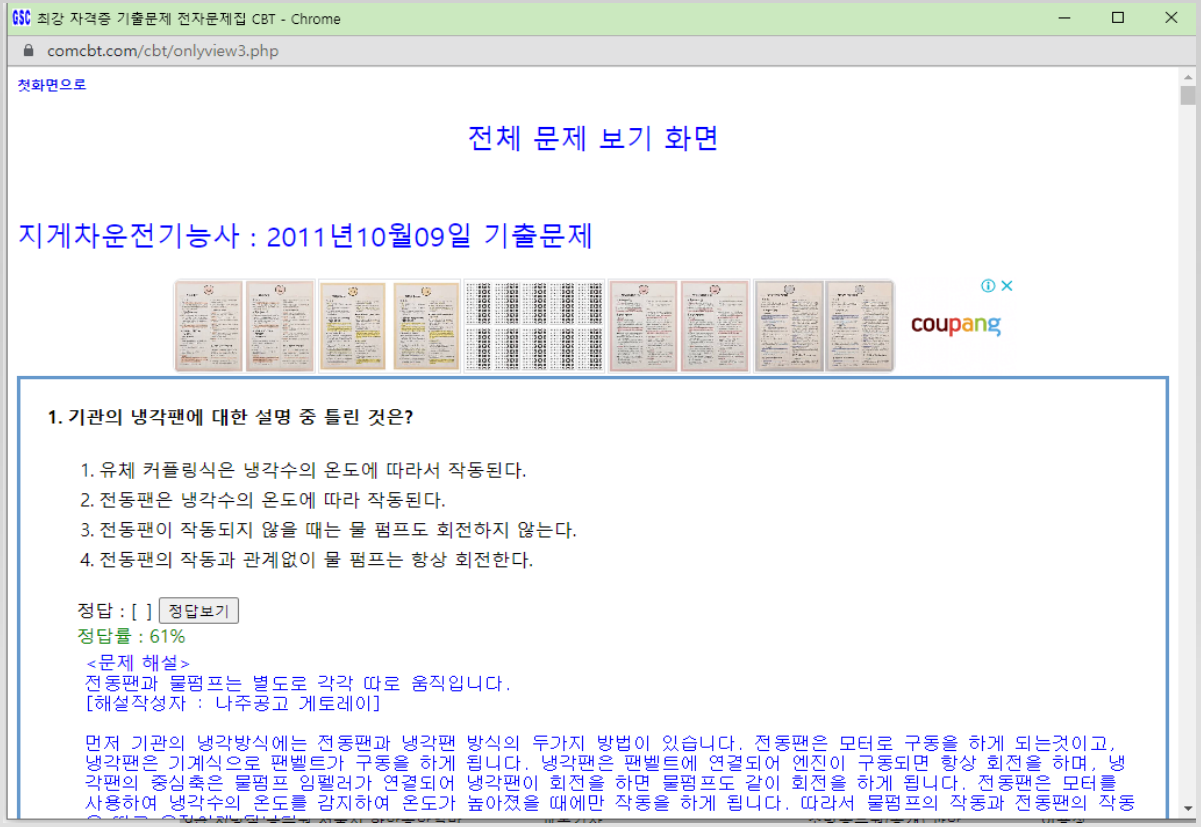
- 9월 27일 ~ 현재 진행중
- 9월 이전에는 소리 관련 프로젝트를 수행 중· 어려움을 느껴 해당 프로젝트를 진행
- 카운트 기반의 문서 표현 및 희소 벡터의 LAS(잠재 의미 분석)차원 축소까지 공부 함. 하지만 LSA 부분은 부족함을 느껴 PPT에서 제외함
- 문맥을 반영하는 RNN과 같은 딥러닝 기술 및 문서 자체를 임베딩 하는 Word2Vec은 아직 시행하지 않음

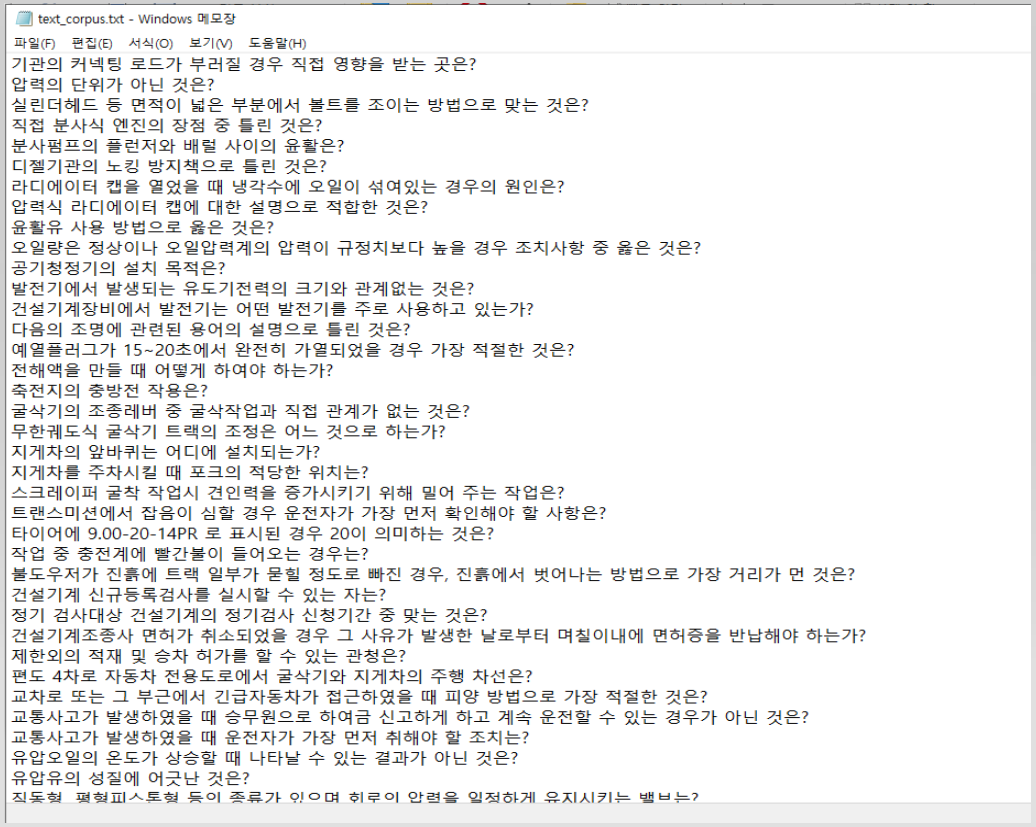
데이터 추출 과정

- 데이터는 04년 ~ 11년 지게차운전기능사(필기) 26개의 시험지 데이터를 구함. 총 1560개의 문서 추출.

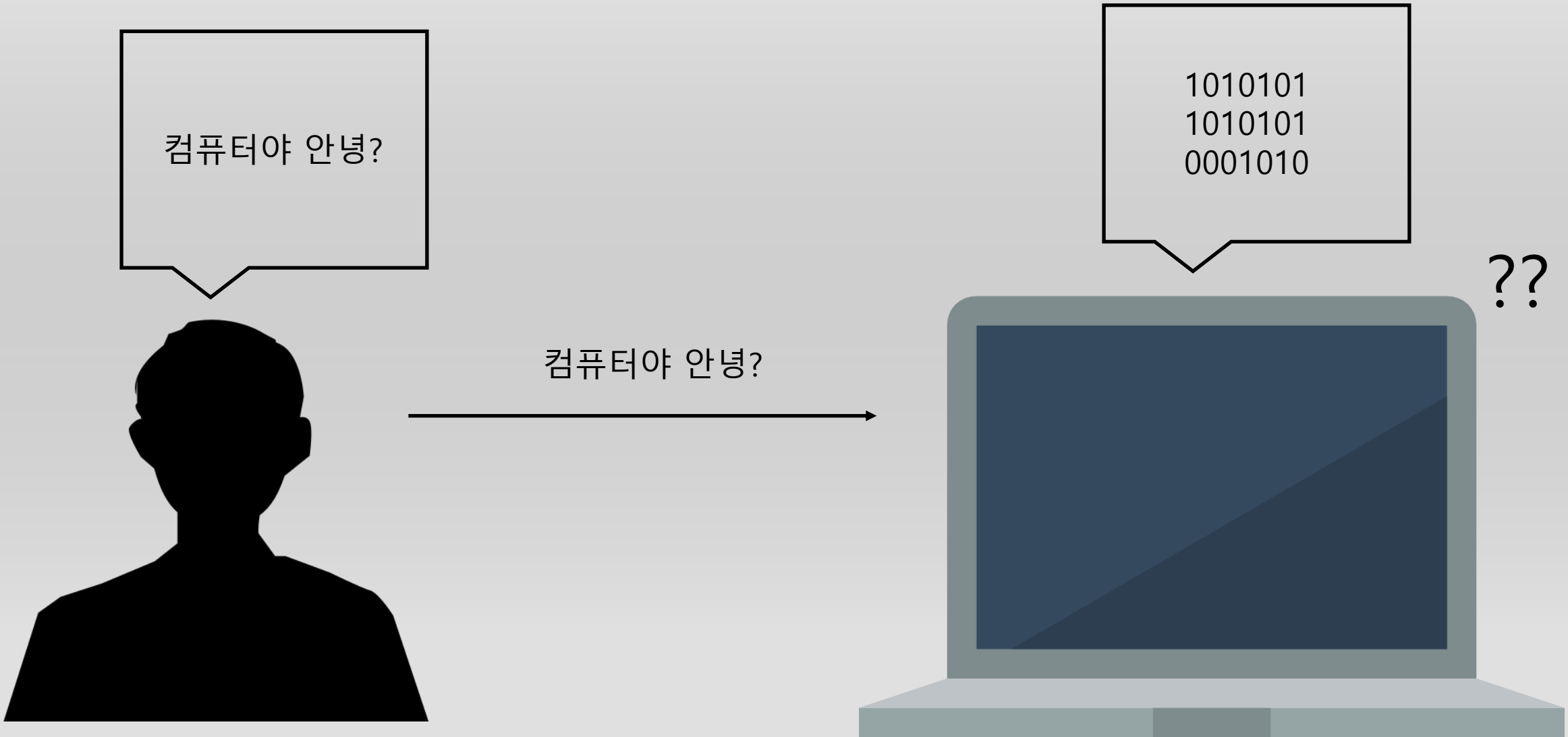
- 자격증 웹페이지에 널리 알려져 있는 cbt웹사이트를 이용

- 문제들을 웹 크롤링을 통해 텍스트 추출하려 했지만 cbt 관리자님이 안된다는 메일을 보내셔서 .hwp 문제집을 하나 하나 다운로드 한 뒤 해당 파일을 html로 변환하고 BeautifulSoup 라이브러리로 크롤링





비정형 데이터의 딜레마



비정형 데이터의 딜레마

[1, 0, 1, 1]

1. 사용자가 원하는 문장을 배열 변환

2. 변환한 배열을 컴퓨터에게 전달

컴퓨터야 안녕?

3. 사용자가 원하는 작업을 컴퓨터가 AI 모델링 기반수행

(해당 예는 질문에 대한 대답을 하는 자연어 처리 모델을 구현했다고 가정)

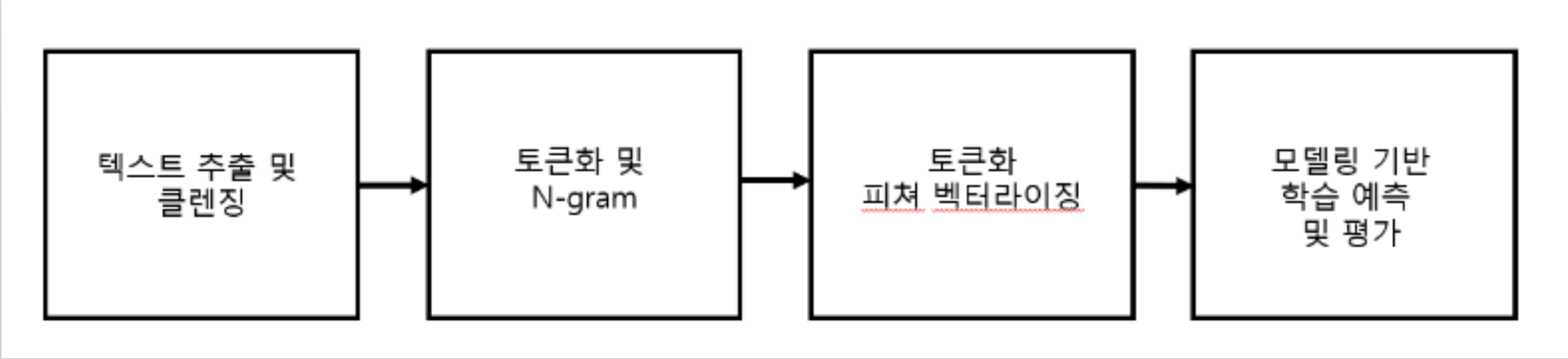
4. 모델에 의해 출력 배열 [0, 1, 1, 0] 산출

[0, 1, 1, 0]

Conversion : '안녕!'

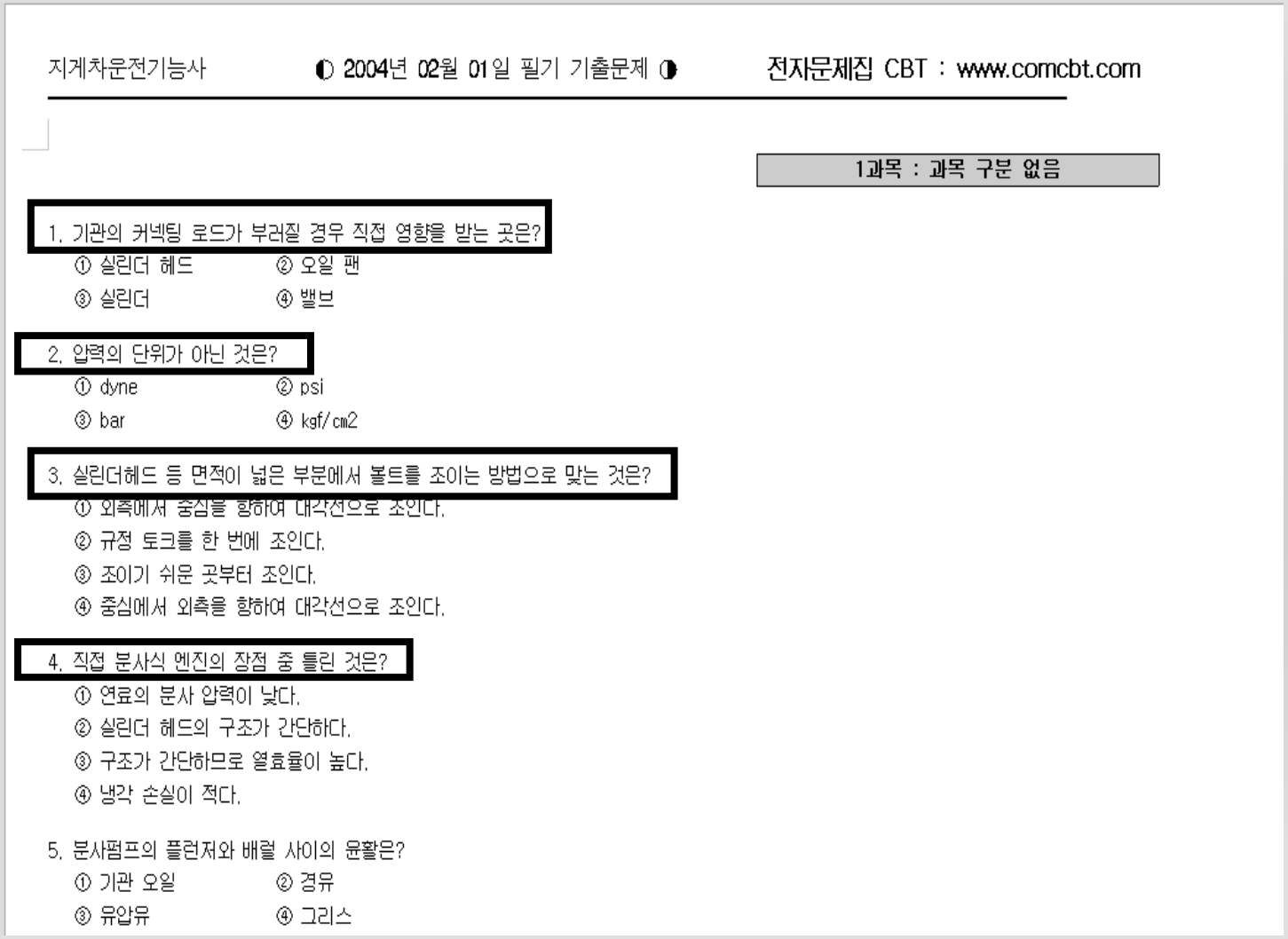
5. 출력 배열을 자연어로 처리하여 전달

자연어 처리의 일반적인 프로세스



텍스트 추출 및 클렌징

- HTML 태그나 기타 노이즈를 제거하는 단계이다.
- 아래의 예를 들면 '1. 기관의 커넥팅 로드가 부러질 경우 직접 영향을 받는 곳은?'에서 1. 은 노이즈에 해당한다.



지게차운전기능사 ◐ 2004년 02월 01일 필기 기출문제 ◑ 전자문제집 CBT : www.comcbt.com

1과목 : 과목 구분 없음

1. 기관의 커넥팅 로드가 부러질 경우 직접 영향을 받는 곳은?
① 실린더 헤드 ② 오일 팬
③ 실린더 ④ 밸브

2. 압력의 단위가 아닌 것은?
① dyne ② psi
③ bar ④ kgf/㎠

3. 실린더헤드 등 면적이 넓은 부분에서 볼트를 조이는 방법으로 맞는 것은?
① 외측에서 중심을 향하여 대각선으로 조인다.
② 규정 토크를 한 번에 조인다.
③ 조이기 쉬운 곳부터 조인다.
④ 중심에서 외측을 향하여 대각선으로 조인다.

4. 직접 분사식 엔진의 장점 중 틀린 것은?
① 연료의 분사 압력이 낮다.
② 실린더 헤드의 구조가 간단하다.
③ 구조가 간단하므로 열효율이 높다.
④ 냉각 손실이 적다.

5. 분사펌프의 플런저와 배럴 사이의 윤활은?
① 기관 오일 ② 경유
③ 유압유 ④ 그리스

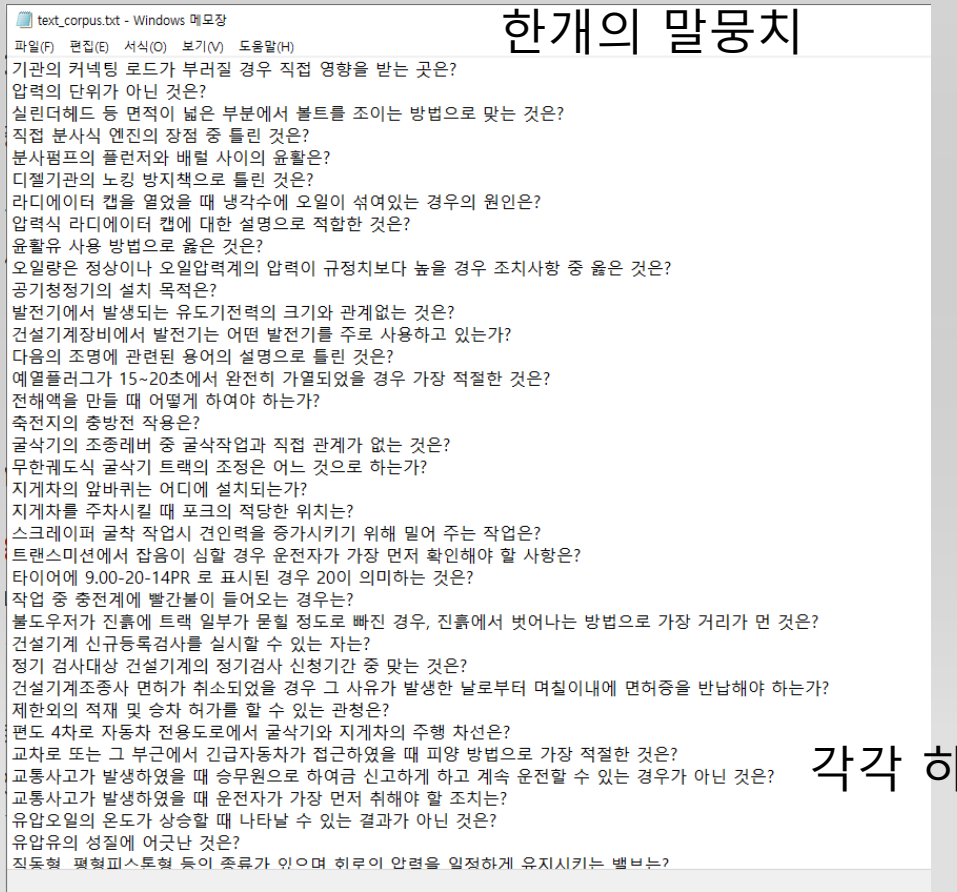
text_corpus.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

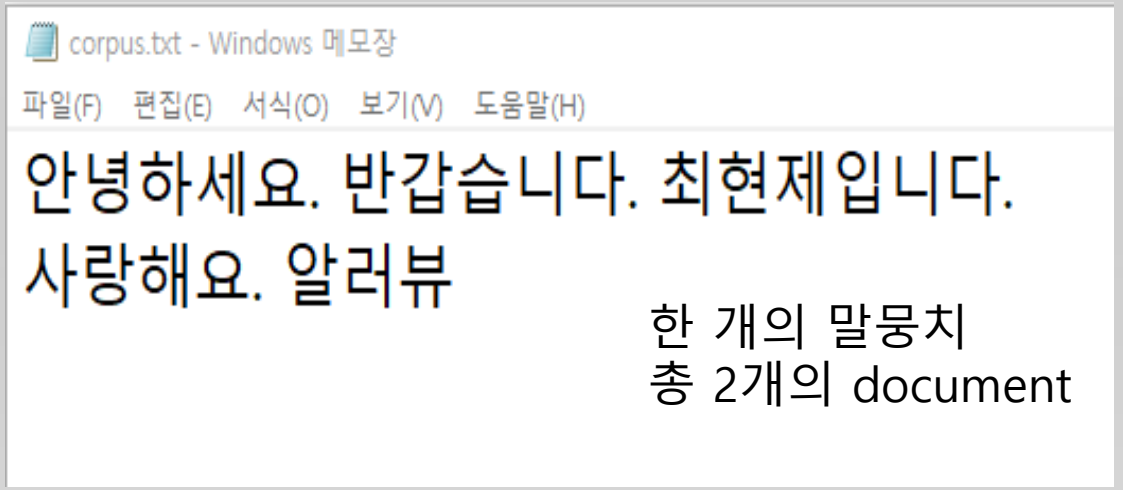
기관의 커넥팅 로드가 부러질 경우 직접 영향을 받는 곳은?
압력의 단위가 아닌 것은?
실린더헤드 등 면적이 넓은 부분에서 볼트를 조이는 방법으로 맞는 것은?
직접 분사식 엔진의 장점 중 틀린 것은?
분사펌프의 플런저와 배럴 사이의 윤활은?
디젤기관의 노킹 방지책으로 틀린 것은?
라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우의 원인은?
압력식 라디에이터 캡에 대한 설명으로 적합한 것은?
윤활유 사용 방법으로 옳은 것은?
오일량은 정상이나 오일압력계의 압력이 규정치보다 높을 경우 조치사항 중 옳은 것은?
공기청정기의 설치 목적은?
발전기에서 발생되는 유도기전력의 크기와 관계없는 것은?
건설기계장비에서 발전기는 어떤 발전기를 주로 사용하고 있는가?
다음의 조명에 관련된 용어의 설명으로 틀린 것은?
예열플러그가 15~20초에서 완전히 가열되었을 경우 가장 적절한 것은?
전해액을 만들 때 어떻게 하여야 하는가?
축전지의 충방전 작용은?
굴삭기의 조종레버 중 굴삭작업과 직접 관계가 없는 것은?
무한궤도식 굴삭기 트랙의 조정은 어느 것으로 하는가?
지게차의 앞바퀴는 어디에 설치되는가?
지게차를 주차시킬 때 포크의 적당한 위치는?
스크레이퍼 굴착 작업시 견인력을 증가시키기 위해 밀어 주는 작업은?
트랜스미션에서 잡음이 심할 경우 운전자가 가장 먼저 확인해야 할 사항은?
타이어에 9.00-20-14PR 로 표시된 경우 20이 의미하는 것은?
작업 중 충전계에 빨간불이 들어오는 경우는?
불도우저가 진흙에 트랙 일부가 묻힐 정도로 빠진 경우, 진흙에서 벗어나는 방법으로 가장 거리가 먼 것은?
건설기계 신규등록검사를 실시할 수 있는 자는?
정기 검사대상 건설기계의 정기검사 신청기간 중 맞는 것은?
건설기계조종사 면허가 취소되었을 경우 그 사유가 발생한 날로부터 며칠이내에 면허증을 반납해야 하는가?
제한외의 적재 및 승차 허가를 할 수 있는 관청은?
편도 4차로 자동차 전용도로에서 굴삭기와 지게차의 주행 차선은?
교차로 또는 그 부근에서 긴급자동차가 접근하였을 때 피양 방법으로 가장 적절한 것은?
교통사고가 발생하였을 때 승무원으로 하여금 신고하게 하고 계속 운전할 수 있는 경우가 아닌 것은?
교통사고가 발생하였을 때 운전자가 가장 먼저 취해야 할 조치는?
유압오일의 온도가 상승할 때 나타날 수 있는 결과가 아닌 것은?
유압유의 성질에 어긋난 것은?
직동형, 평형피스톤형 등의 종류가 있으며 회로의 압력을 일정하게 유지시키는 밸브는?

잠깐 용어정리

- 자연어 처리에서는 말뭉치(Corpus)와 문서(Document) 라는 키워드가 있다.
- 말뭉치(Corpus)란 언어 관련 분야에서 데이터셋이나 테스트 셋을 뜻한다.
- 문서(Document)란 하나의 데이터 단위이며, 문장을 하나 또는 그 이상 포함한다.



한개의 말뭉치



한 개의 말뭉치
총 2개의 document

각각 하나의 문장이 하나의 document

토큰화

- 말뭉치(corpus)에서 문장마다 토큰이라 불리는 단위로 나누는 작업이다.

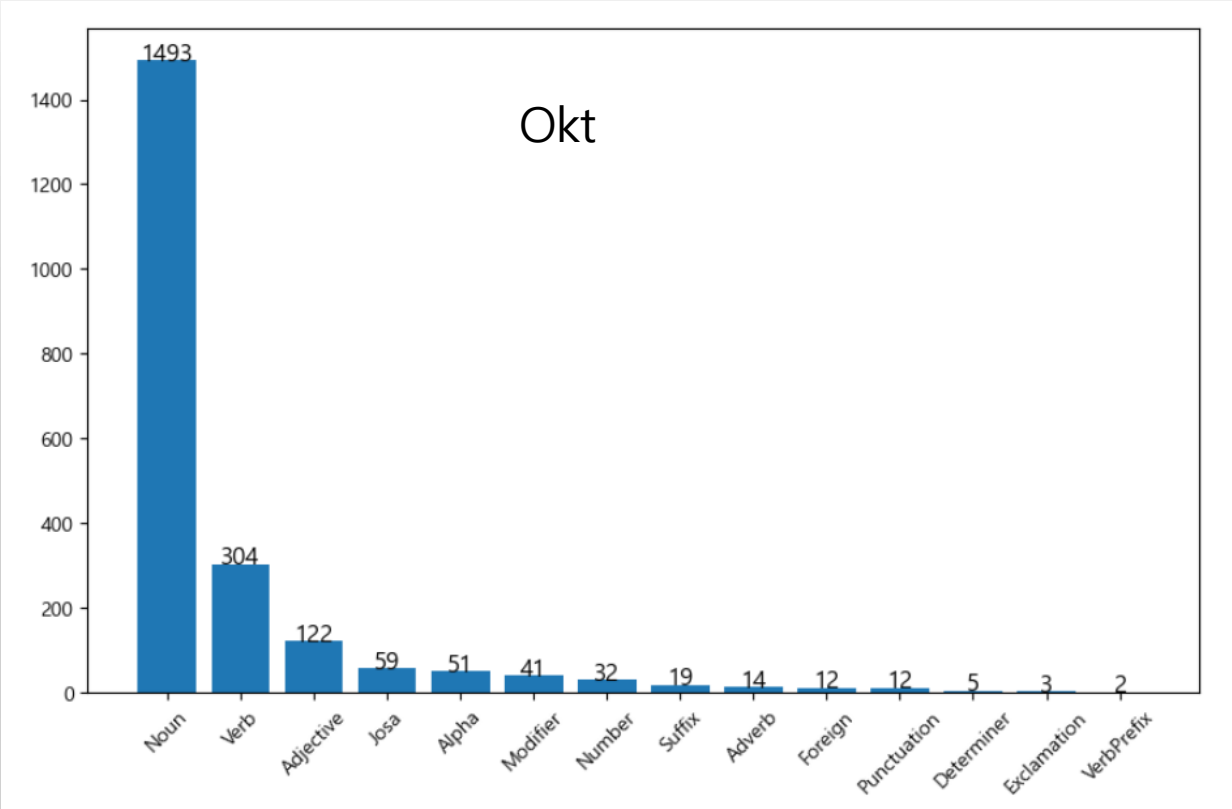
- 문장 토큰화, 단어 토큰화가 있다. 한국어는 단어 토큰화, 그 중에서 의미를 가지는 요소의 가장 작은 단위인 형태소 기준으로 토큰화를 한다.

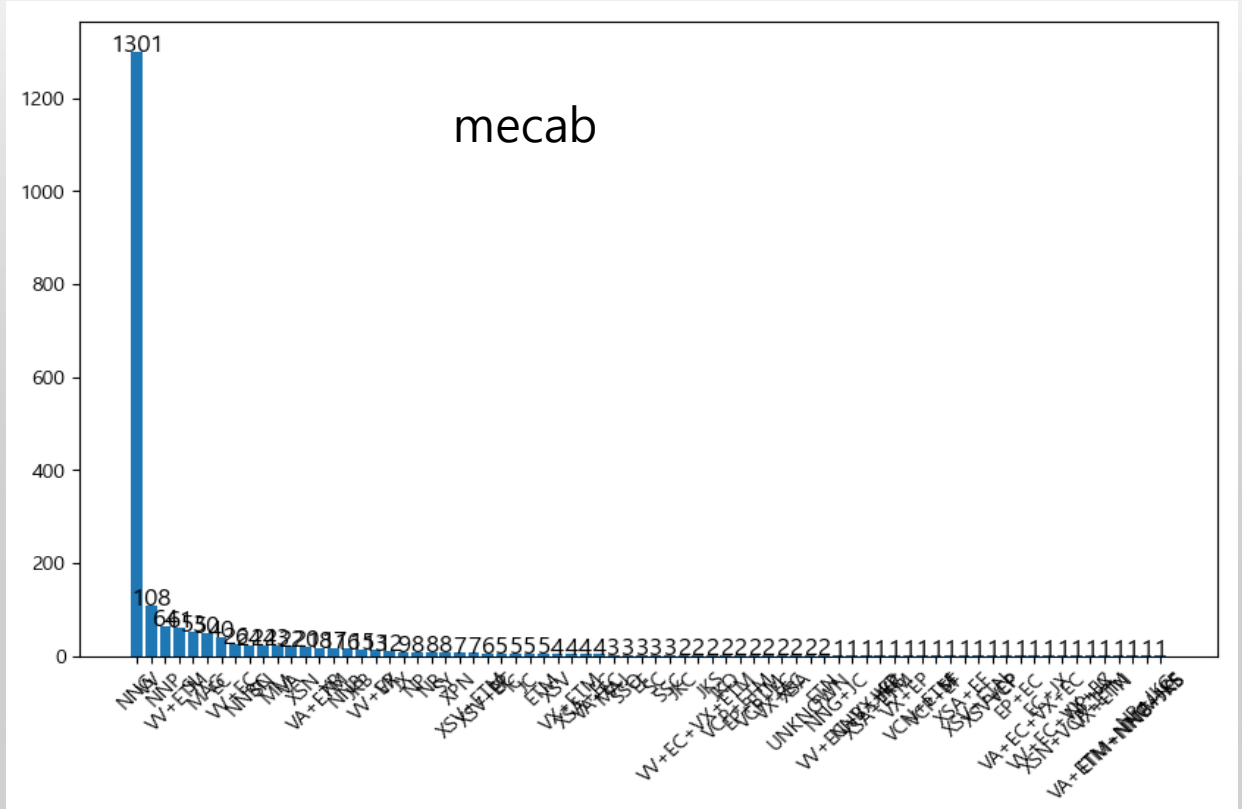
- 파이썬에서는 영어는 nltk를 쓰며, 한글은 konlpy(코엔엘파이) 모듈에서 5개의 형태소 분석기 중 하나를 쓴다.

executed in 1m 9.72s, finished 00:24:43 2022-10-12

```
hannanum - 1560개 문서를 계산할 때 시간 :  5.267852783203125 sec
komoran - 1560개 문서를 계산할 때 시간 :  1.9653270244598389 sec
okt - 1560개 문서를 계산할 때 시간 :  4.080617427825928 sec
mecab - 1560개 문서를 계산할 때 시간 :  1.143613338470459 sec
kkma - 1560개 문서를 계산할 때 시간 :  57.239693880081118 sec
```

토큰화 (Okt vs mecab)

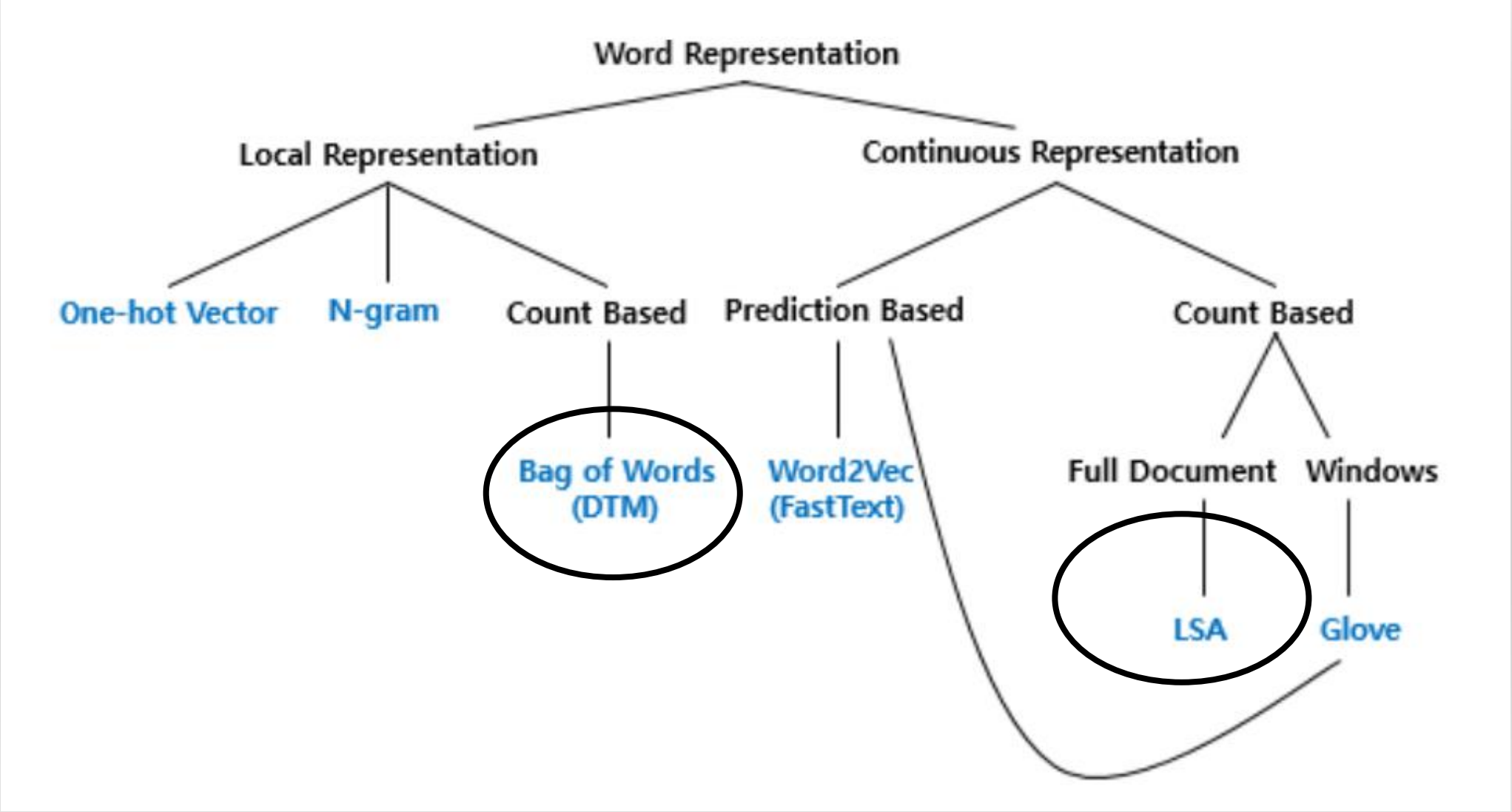




Okt는 형태소 분류기는 품사를 총 14개로 분류

Mecab 형태소 분류기는 독보적으로
문서의 품사를 세세하게 분류

토큰화 피쳐 벡터라이징
- 토큰화 한 기준으로 벡터화 하는 것



토큰화 피쳐 벡터라이징
- CountVectorize

문서
'직접 분사식 엔진 의 장점 중 틀린 것 은 ?'
'디젤 기관 의 노킹 방지책 으로 틀린 것 은 ?'
'압력 식 라디에이터 캡 에 대한 설명 으로 적합 한 것 은 ?'

{'직접': 13, '분사식': 6, '엔진': 9, '장점': 11, '틀린': 14, '디젤': 3, '기관': 0, '노킹': 1, '방지책': 5, '으로': 10, '압력': 8, '라디에이터': 4, '대한': 2, '설명': 7, '적합': 12}

기관	노킹	대한	디젤	라디에이터	방지책	분사식	설명	압력	엔진	으로	장점	적합	직접	틀린
0	0	0	0	0	0	1	0	0	1	0	1	0	1	1
1	1	0	1	0	1	0	0	0	0	1	0	0	0	1
0	0	1	0	1	0	0	1	1	0	1	0	1	0	0

토큰화 피쳐 벡터라이징
- TF-IDF

TF : 전체 문서에서 특정 단어가 나온 횟수

IDF : $\ln\left(\frac{1+n}{1+df}\right)+1$ (n은 말뭉치 내 문서의 총 개수, df는 말뭉치 내 문서 개수에서 특정 단어가 등장한 빈도

각각의 문서에서 자주 등장하는 단어에 높은 가중치를 주고, 모든 문서에서 자주 등장하는 단어에 페널티를 준다.

문서
'엄마 나 닭도리탕 해 주 세요 . 닭도리탕 !'
'엄마 사랑 해 .'

- 해당 말뭉치에서 '닭도리탕' 이라는 단어의 df는 2가 아닌 1이다.
- Why? 두 문서 중 한 문서에 '닭도리탕' 이라는 단어가 있기 때문이다.

토큰화 피쳐 벡터라이징
- TF-IDF 예시

기관	노킹	대한	디젤	라디에이터	방지책	분사식	설명	압력	엔진	으로	장점	적합	직접	틀린
0	0	0	0	0	0	1	0	0	1	0	1	0	1	1
1	1	0	1	0	1	0	0	0	0	1	0	0	0	1
0	0	1	0	1	0	0	1	1	0	1	0	1	0	0

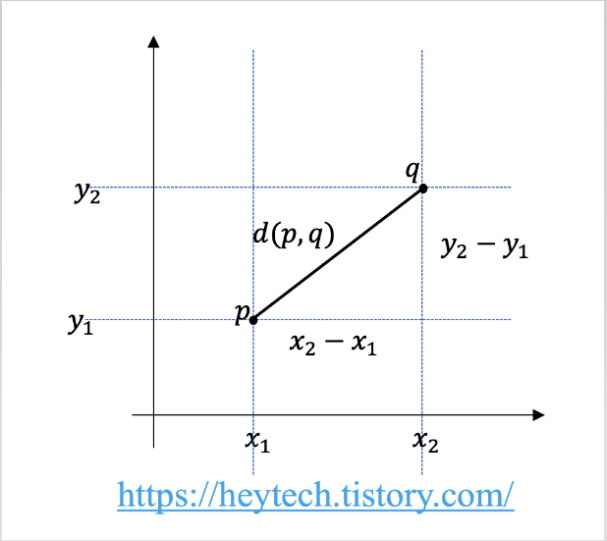
$\text{IDF}: \ln\left(\frac{1+n}{1+\text{d}f}\right)+1$, L2 norm을 시행하여 정규화한다.

기관	노킹	대한	디젤	라디에이터	방지책	분사식	설명	압력	엔진	으로	장점	적합	직접	틀린
0	0	0	0	0	0	0.46	0	0	0.46	0	0.46	0	0.46	0.35
0.44	0.44	0	0.44	0	0.44	0	0	0	0	0.33	0	0	0	0.33
0	0	0.42	0	0.42	0	0	0.42	0.42	0	0.32	0	0.42	0	0

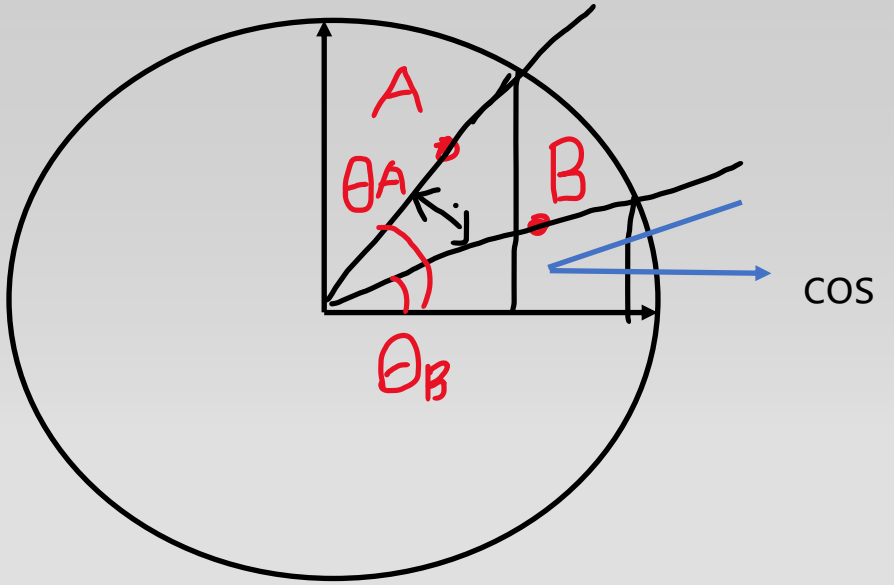
유사도 평가
- 유클리디안 거리 vs 코사인 유사도 (2차원 벡터라고 가정)

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

크기에 따른 유사도 측정



방향에 따른 유사도 측정



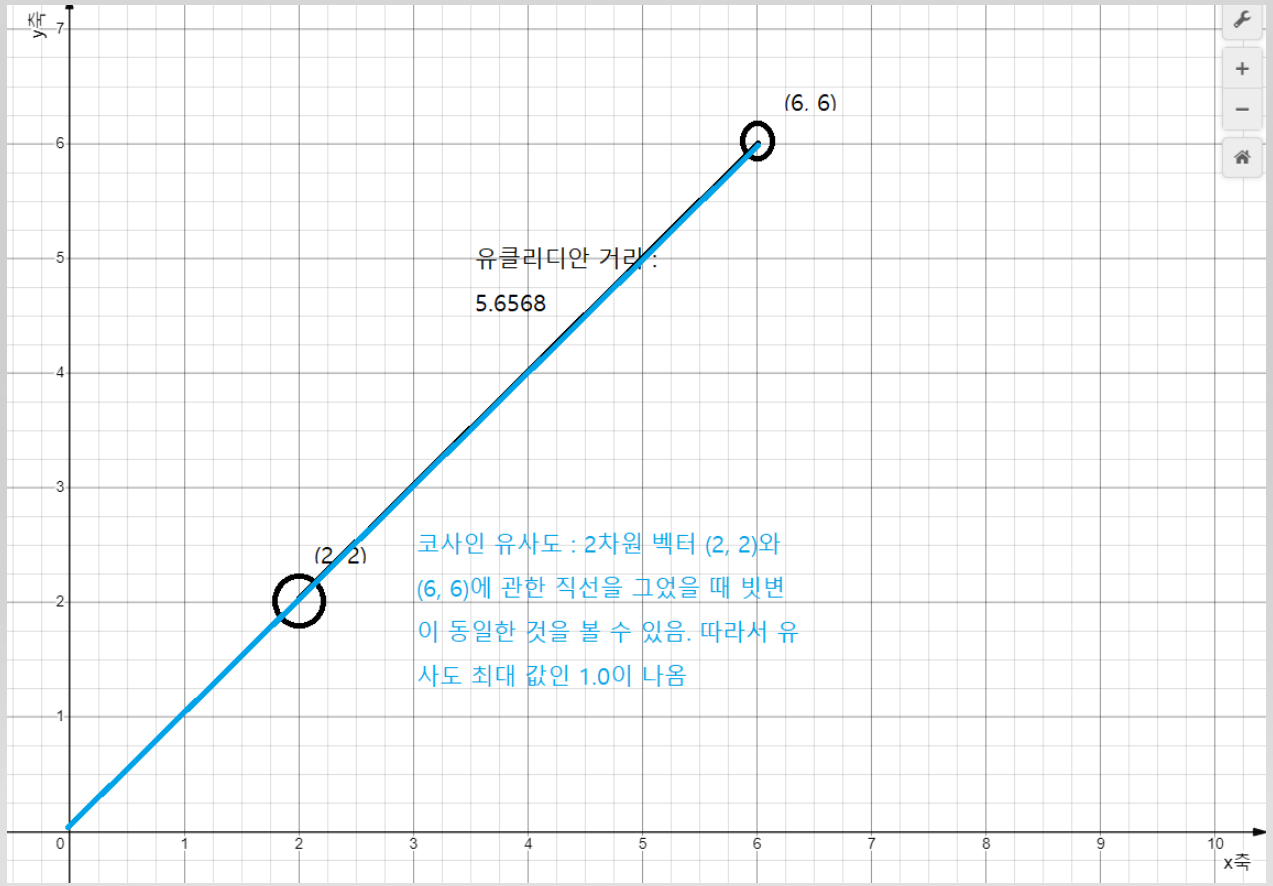
예시 [토끼 : 0], [거북이 : 1]

Document1 : 6, 6

Document2 : 2, 2

```
In [351]: euclidean_distances([[6, 6]], [[2, 2]])

          executed in 23ms, finished 00:36:33 2022-10-19

Out[351]: array([[5.65685425]])
```

```
In [352]: cosine_similarity([[6, 6]], [[2, 2]])

          executed in 22ms, finished 00:36:50 2022-10-19

Out[352]: array([[1.]])
```

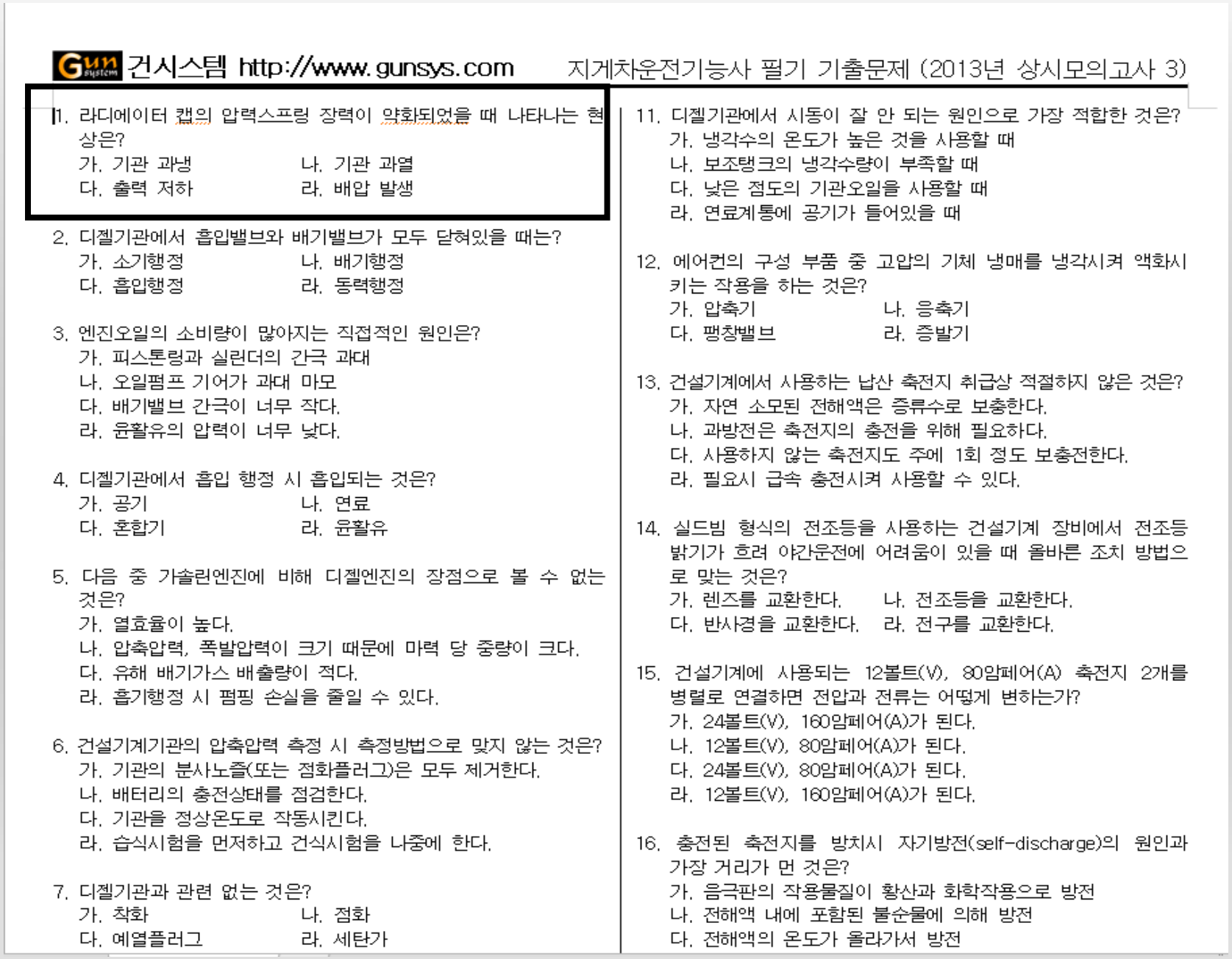


전처리
실습

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'

1. 형태소 분석기 : Okt
2. 임베딩 : Tfidf
3. 유사도 분석

- 코사인 유사도, 유클리디안



건시스템 http://www.gunsys.com 지게차운전기능사 필기 기출문제 (2013년 상시모의고사 3)

1. 라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?
가. 기관 과냉 나. 기관 과열
다. 출력 저하 라. 배압 발생

2. 디젤기관에서 흡입밸브와 배기밸브가 모두 닫혀있을 때는?
가. 소기행정 나. 배기행정
다. 흡입행정 라. 동력행정

3. 엔진오일의 소비량이 많아지는 직접적인 원인은?
가. 피스톤링과 실린더의 간극 과대
나. 오일펌프 기어가 과대 마모
다. 배기밸브 간극이 너무 작다.
라. 윤활유의 압력이 너무 낮다.

4. 디젤기관에서 흡입 행정 시 흡입되는 것은?
가. 공기 나. 연료
다. 혼합기 라. 윤활유

5. 다음 중 가솔린엔진에 비해 디젤엔진의 장점으로 볼 수 없는 것은?
가. 열효율이 높다.
나. 압축압력, 폭발압력이 크기 때문에 마력 당 중량이 크다.
다. 유해 배기가스 배출량이 적다.
라. 흡기행정 시 펌핑 손실을 줄일 수 있다.

6. 건설기계기관의 압축압력 측정 시 측정방법으로 맞지 않는 것은?
가. 기관의 분사노즐(또는 점화플러그)은 모두 제거한다.
나. 배터리의 충전상태를 점검한다.
다. 기관을 정상온도로 작동시킨다.
라. 습식시험을 먼저하고 건식시험을 나중에 한다.

7. 디젤기관과 관련 없는 것은?
가. 착화 나. 점화
다. 예열플러그 라. 세탄가

11. 디젤기관에서 시동이 잘 안 되는 원인으로 가장 적합한 것은?
가. 냉각수의 온도가 높은 것을 사용할 때
나. 보조탱크의 냉각수량이 부족할 때
다. 낮은 점도의 기관오일을 사용할 때
라. 연료계통에 공기가 들어있을 때

12. 에어컨의 구성 부품 중 고압의 기체 냉매를 냉각시켜 액화시키는 작용을 하는 것은?
가. 압축기 나. 응축기
다. 팽창밸브 라. 증발기

13. 건설기계에서 사용하는 납산 축전지 취급상 적절하지 않은 것은?
가. 자연 소모된 전해액은 증류수로 보충한다.
나. 과방전은 축전지의 충전을 위해 필요하다.
다. 사용하지 않는 축전지도 주에 1회 정도 보충전한다.
라. 필요시 급속 충전시켜 사용할 수 있다.

14. 실드빔 형식의 전조등을 사용하는 건설기계 장비에서 전조등 밝기가 흐려 야간운전에 어려움이 있을 때 올바른 조치 방법으로 맞는 것은?
가. 렌즈를 교환한다. 나. 전조등을 교환한다.
다. 반사경을 교환한다. 라. 전구를 교환한다.

15. 건설기계에 사용되는 12볼트(V), 80암페어(A) 축전지 2개를 병렬로 연결하면 전압과 전류는 어떻게 변하는가?
가. 24볼트(V), 160암페어(A)가 된다.
나. 12볼트(V), 80암페어(A)가 된다.
다. 24볼트(V), 80암페어(A)가 된다.
라. 12볼트(V), 160암페어(A)가 된다.

16. 충전된 축전지를 방치시 자기방전(self-discharge)의 원인과 가장 거리가 먼 것은?
가. 음극판의 작용물질이 황산과 화학작용으로 방전
나. 전해액 내에 포함된 불순물에 의해 방전
다. 전해액의 온도가 올라가서 방전

Vectorize_sim이라는 함수를 만들어 okt -> tfidf -> 코사인 유사도, 유클리디안 거리를 데이터프레임으로 만드는 과정을 구현

Out [220] :

	코사인 유사도	문서	유클리디안 거리	문서
1263	0.596285	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?	2.449490	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
157	0.507093	유압조정 밸브에서 조정 스프링의 장력이 를 때 나타나는 현상은?	2.449490	유압조정 밸브에서 조정 스프링의 장력이 를 때 나타나는 현상은?
461	0.507093	유압조정 밸브에서 조정 스프링의 장력이 를 때 나타나는 현상은?	2.449490	유압조정 밸브에서 조정 스프링의 장력이 를 때 나타나는 현상은?
1086	0.447214	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?	2.828427	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
66	0.338062	팬벨트의 장력이 너무 강할 경우에 발생되는 현상은?	2.828427	팬벨트의 장력이 너무 강할 경우에 발생되는 현상은?
...	...	...	...	...
526	0.000000	렌치 작업시의 주의사항 설명 중 틀린 것은?	3.316625	렌치 작업시의 주의사항 설명 중 틀린 것은?
524	0.000000	유압유 취급에 대한 설명으로 틀린 것은?	3.316625	유압유 취급에 대한 설명으로 틀린 것은?
523	0.000000	유압 액츄에이터(작업장치)를 교환하였을 경우 반드시 해야 할 작업이 아닌 것은?	3.872983	유압 액츄에이터(작업장치)를 교환하였을 경우 반드시 해야 할 작업이 아닌 것은?
522	0.000000	회로 내 유체의 흐르는 방향을 조절하는데 쓰이는 밸브는?	3.464102	회로 내 유체의 흐르는 방향을 조절하는데 쓰이는 밸브는?
1559	0.000000	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?	3.605551	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?

1560 rows × 4 columns

1. 문서를 어절(띄어쓰기) 별로 그냥 TfidfVectorizer를 해볼까?
- 영어는 단어 자체로 사전이 만들어져도 이상 없는데 한글은 조사가 있어서 유사도 성능도 떨어지는 것 같다.

```
DFM : (1560, 3750)
['00' '100m' '100만' '100만원의' '100미터' '100분의' '100을' '10개' '10개로' '10만원마다'
 '12v' '12v이고' '12ⅴ' '14pr' '15' '154kv' '1개' '1년간' '1사이클을' '1인당'
 '1제곱센티미터' '1종' '1톤' '20' '20m' '200이' '20초에서' '22' '24v' '280' '2800이면'
 '2개' '2개를' '2명' '2종류의' '30m' '30미터' '3m' '3가지' '3대' '3대작용이' '3요소에' '3요소와'
 '3톤' '4000rpm일' '4m' '4m이상' '4기통기관' '4차로' '4차로의' '4행정' '4행정기관에서' '4행정으로'
 '50을' '5kw' '5mpa이다' '5년' '5명의' '6각' '6기통' '6기통기관이' '6옴일' '8m' '8m미만인'
 '9kv' 'ac' 'actuator' 'alternator' 'arm' 'a가' 'a의' 'charger' 'cm' 'cm로'
 'comp' 'coupling' 'crank' 'cylinder' 'de' 'driver' 'fire' 'flow' 'fluid'
 'free' 'full' 'gpm의' 'gpm이' 'grader' 'grow' 'ilo' 'key' 'kv' 'lng' 'low'
 'maintenance' 'mesh' 'mf' 'miss' 'motor' 'mpa' 'm마다' 'm의' 'm이내를' 'm이상의'
 'm이상인가' 'nox' 'on' 'plp관' 'pressure' 'rpm' 'rpm이란' 'shaft' 'st' 'surge'
 'turbo' '⊖방향을' '가고' '가공' '가공송전선로' '가까이' '가는' '가는데' '가능' '가능한' '가동하고' '가득'
 '가변' '가변용량' '가속되는' '가속페달의' '가스' '가스가' '가스공급시설을' '가스누출' '가스는' '가스도매사업자의'
 '가스를' '가스배관' '가스배관과' '가스배관과의' '가스배관을' '가스배관의' '가스배관이' '가스보호포가' '가스사용자가'
 '가스안전' '가스안전영향평가서를' '가스용접에서' '가스장치의' '가시거리가' '가압식' '가압하여' '가연성' '가열되었을'
 '가이드링의' '가장' '가장거리가' '가장자리에' '가져야' '가지' '가지고' '가진' '가파른' '가하면' '가해진'
 '가했을' '각각' '각도는' '각종' '간격으로' '간극은' '간극을' '간극이' '감소' '감소되었을' '감속' '감속기의'
 '감속으로' '감속작용을' '감시하고' '감전되거나' '감전사고의' '감전이' '감전재해' '감전재해의' '감자기' '감산'
```

	코사인 유사도	문서	유클리디안 거리	문서
1263	0.601678	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?	0.892549	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
1086	0.458221	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?	1.040941	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
157	0.406322	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?	1.089659	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
461	0.406322	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?	1.089659	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
66	0.279658	팬벨트의 장력이 너무 강할 경우에 발생되는 현상은?	1.200285	팬벨트의 장력이 너무 강할 경우에 발생되는 현상은?
...	...	...	...	...
526	0.000000	렌치 작업시의 주의사항 설명 중 틀린 것은?	1.414214	렌치 작업시의 주의사항 설명 중 틀린 것은?

2. 그러면 형태소 기준으로 문장을 띄어 쓰기 후 TfidfVectorizer를 해야겠다.
- 좌측에 있는 어절 별 띄어쓰기를 우측인 형태소 기준으로 띄어쓰기 한 문서 리스트

```
['기관의 커넥팅 로드가 부러질 경우 직접 영향을 받는 곳은?',
 '압력의 단위가 아닌 것은?',
 '실린더헤드 등 면적이 넓은 부분에서 볼트를 조이는 방법으로 맞는 것은?',
 '직접 분사식 엔진의 장점 중 틀린 것은?',
 '분사펌프의 플런저와 배럴 사이의 윤활은?',
 '디젤기관의 노킹 방지책으로 틀린 것은?',
 '라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우의 원인은?',
 '압력식 라디에이터 캡에 대한 설명으로 적합한 것은?',
 '윤활유 사용 방법으로 옳은 것은?',
 '오일량은 정상이나 오일압력계의 압력이 규정치보다 높을 경우 조치사항 중 옳은 것은?',
 '공기청정기의 설치 목적은?',
 '발전기에서 발생되는 유도기전력의 크기와 관계없는 것은?',
 '건설기계장비에서 발전기는 어떤 발전기를 주로 사용하고 있는가?',
 '다음의 조명에 관련된 용어의 설명으로 틀린 것은?',
 '예열플러그가 15~20초에서 완전히 가열되었을 경우 가장 적절한 것은?',
 '전해액을 만들 때 어떻게 하여야 하는가?',
 '축전지의 충방전 작용은?',
 '굴삭기의 조종레버 중 굴삭작업과 직접 관계가 없는 것은?',
 '무한궤도식 굴삭기 트랙의 조정은 어느 것으로 하는가?',
```

```
['기관 의 커넥팅 로드 가 부러질 경우 직접 영향 을 받는 곳 은 ?',
 '압력 의 단위 가 아닌 것 은 ?',
 '실린더헤드 등 면적 이 넓은 부분 에서 볼트 를 조이는 방법 으로 맞는 것 은 ?',
 '직접 분사 식 엔진 의 장점 중 틀린 것 은 ?',
 '분사 펌프 의 플런저 와 배럴 사이 의 윤활 은 ?',
 '디젤기관 의 노킹 방지책 으로 틀린 것 은 ?',
 '라디에이터 캡 을 열었을 때 냉각수 에 오일이 섞여있는 경우 의 원인 은 ?',
 '압력 식 라디에이터 캡 에 대한 설명 으로 적합한 것 은 ?',
 '윤활유 사용 방법 으로 옳은 것 은 ?',
 '오 일 량 은 정상 이나 오 일 압력계 의 압력 이 규 정치 보다 높을 경우 조치 사항 중 옳은 것 은 ?',
 '공기청정기 의 설치 목적 은 ?',
 '발전기 에서 발생 되는 유도기 전력 의 크기 와 관계 없는 것 은 ?',
 '건설 기계 장비 에서 발전기 는 어떤 발전기 를 주로 사용 하고 있는가 ?',
 '다음 의 조명 에 관련 된 용어 의 설명 으로 틀린 것 은 ?',
 '예 열 플러그 가 15 ~ 20초 에서 완전히 가열 되었을 경우 가장 적절한 것 은 ?',
 '전해 액 을 만들 때 어떻게 하여야 하는가 ?',
 '축전지 의 충 방전 작용 은 ?',
 '굴삭기 의 조종 레버 중 굴삭 작업 과 직접 관계 가 없는 것 은 ?',
 '무한궤도 식 굴삭기 트랙 의 조정은 어느 것 으로 하는가 ?',
```

2. 그러면 형태소 기준으로 문장을 띄어 쓰기 후 TfidfVectorizer를 해야겠다.
- 테스트 문서인 '라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?' 과 가장 유사한 문서는 다음과 같다. 같은 문장인데 이렇게 유사도가 다르다고?! 뭔가 이상하다.

	코사인 유사도	문서	유클리디안 거리	문서
1263	0.603189	라디에이터 캡 의 스프링 이 파손 되었을 때 가장 먼저 나타나는 현상 은 ?	0.890855	라디에이터 캡 의 스프링 이 파손 되었을 때 가장 먼저 나타나는 현상 은 ?
1242	0.526060	유압 계통 에서 릴리 프 밸브 스프링 의 장력 이 약화 될 때 발생 될 수 있는 현...	0.973591	유압 계통 에서 릴리 프 밸브 스프링 의 장력 이 약화 될 때 발생 될 수 있는 현...
1086	0.449622	라디에이터 캡 의 스프링 이 파손 되었을 때 가장 먼저 나타나 는 현상 은 ?	1.049169	라디에이터 캡 의 스프링 이 파손 되었을 때 가장 먼저 나타나 는 현상 은 ?
525	0.445083	유압펌프 의 압력 조절 밸브 스프링 장력 이 높은 것 을 사용 하면 나타나는 현상 ...	1.053486	유압펌프 의 압력 조절 밸브 스프링 장력 이 높은 것 을 사용 하면 나타나는 현상 ...
157	0.435442	유압 조정 밸브 에서 조정 스프링 의 장력 이 클 때 나타나는 현상 은 ?	1.062599	유압 조정 밸브 에서 조정 스프링 의 장력 이 클 때 나타나는 현상 은 ?
...	...	...	...	...
551	0.000000	겨울철 에 기동전 동기 크 랭킹 회전 수가 낮아지는 원인 이 아닌 것 은 ?	1.414214	겨울철 에 기동전 동기 크 랭킹 회전 수가 낮아지는 원인 이 아닌 것 은 ?
550	0.000000	예 연 소실 식 연 소실 에 대한 설명 으로 가장 거리 가 먼 것 은 ?	1.414214	예 연 소실 식 연 소실 에 대한 설명 으로 가장 거리 가 먼 것 은 ?
549	0.000000	엔진오일 에 대한 설명 으로 맞는 것 은 ?	1.414214	엔진오일 에 대한 설명 으로 맞는 것 은 ?
547	0.000000	디젤엔진 의 진동 원인 이 아닌 것 은 ?	1.414214	디젤엔진 의 진동 원인 이 아닌 것 은 ?
1559	0.000000	굴착 도중 전력 케이블 표지 시트 가 나왔을 경우 의 조치 사항 으로 적합한 것 은 ?	1.414214	굴착 도중 전력 케이블 표지 시트 가 나왔을 경우 의 조치 사항 으로 적합한 것 은 ?

1560 rows × 4 columns

주요 원인 : 말뭉치에 있는 문서들을 클렌징을 잘 하지 못해 아래의 '는'이 형태소로 따로 추출 됨

라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?

In [371]:
```
okt.pos('라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?')
```

executed in 21ms, finished 02:34:11 2022-10-19

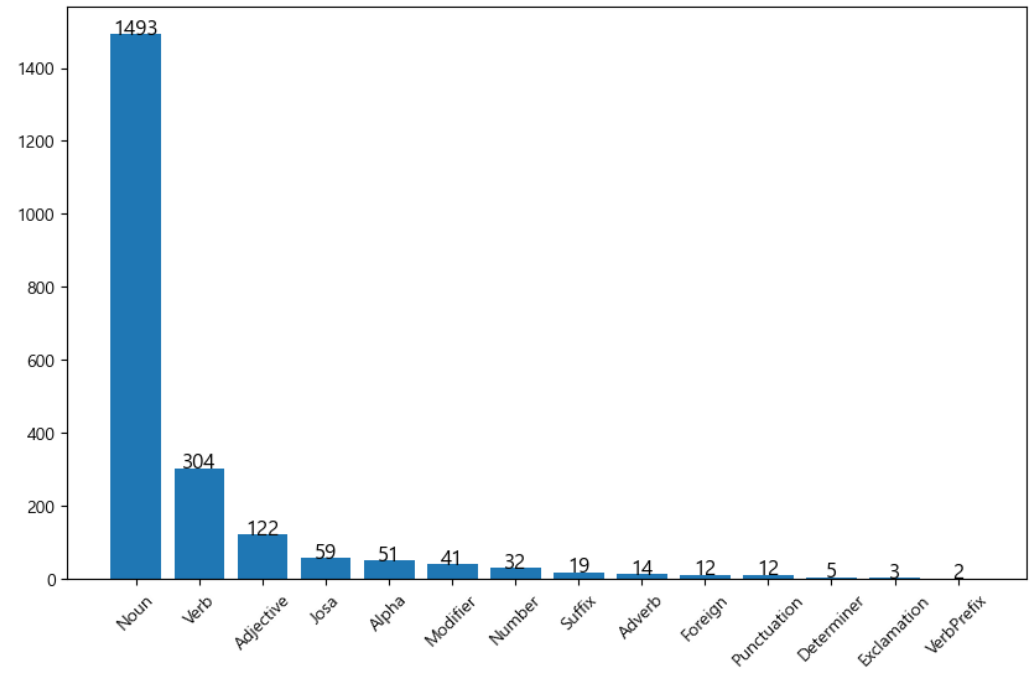
Out[371]:
```
[('라디에이터', 'Noun'),
 ('캡', 'Noun'),
 ('의', 'Josa'),
 ('스프링', 'Noun'),
 ('이', 'Josa'),
 ('파손', 'Noun'),
 ('되었을', 'Verb'),
 ('때', 'Noun'),
 ('가장', 'Noun'),
 ('먼저', 'Noun'),
 ('나타나는', 'Verb'),
 ('현상', 'Noun'),
 ('은', 'Josa'),
 ('?', 'Punctuation')]
```

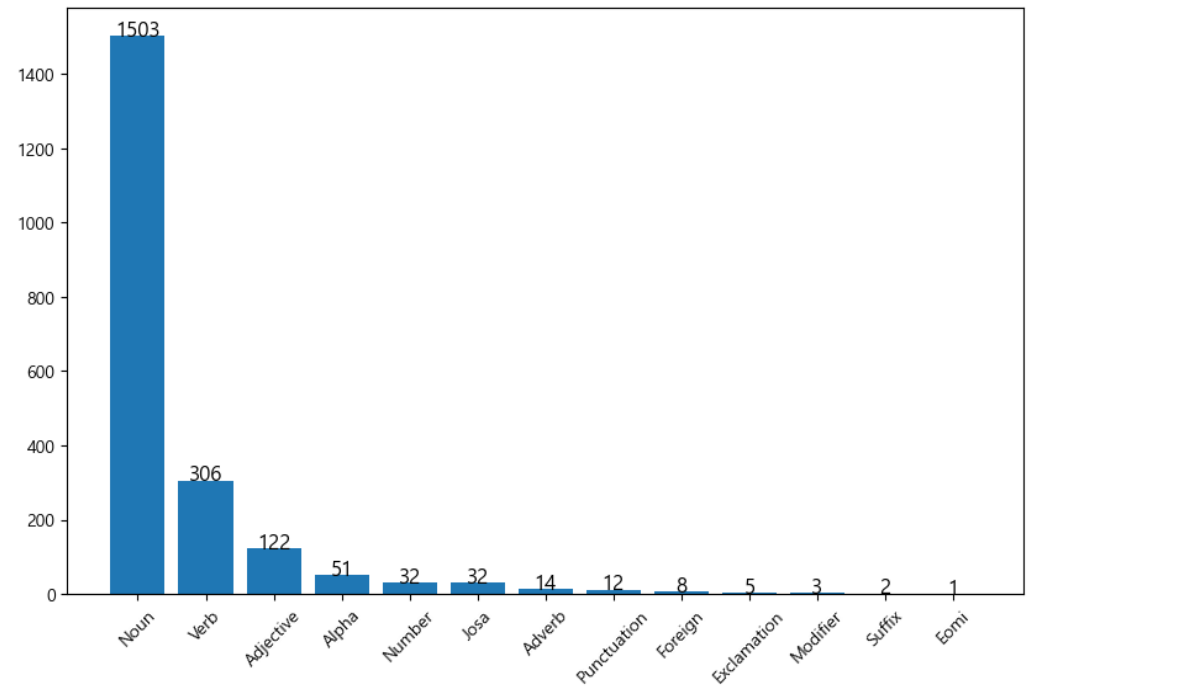
In [369]:
```
okt.pos('라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?')
```

executed in 13ms, finished 02:32:24 2022-10-19

Out[369]:
```
[('라디에이터', 'Noun'),
 ('캡', 'Noun'),
 ('의', 'Josa'),
 ('스프링', 'Noun'),
 ('이', 'Josa'),
 ('파손', 'Noun'),
 ('되었을', 'Verb'),
 ('때', 'Noun'),
 ('가장', 'Noun'),
 ('먼저', 'Noun'),
 ('나타나', 'Verb'),
 ('는', 'Verb'),
 ('현상', 'Noun'),
 ('은', 'Josa'),
 ('?', 'Punctuation')]
```

3. 그러면 형태소 기준으로 문장을 띄어 쓰기 한 데이터를 명사만 추출해서 TfidfVectorizer를 해야겠다.
- 그런데 어절 별로 띄어쓰기 된 원본 데이터와 형태소 별로 띄어쓰기 된 데이터의 명사 개수가 다르다?





원본 데이터

형태소로 띄어쓰기 된 데이터

형태소 데이터를 기준으로 명사 개수가 확 늘었고 조사 개수가 줄어들었다?

아! 형태소 기준으로 띄어쓰기를 하면 okt 형태소 분석기가 일부 조사를 명사로 인식하여 문제가 생김

'다음 중 교류 발전기 의 부품 이 아닌 것 은 ?',
'건설 기계 장비 가 시동 이 되지 않아 시 동 장치 를 점검 하고 있다',
'트랙 장치 에서 트랙 과 아이 들러의 충격 을 완화 시키기 위해 설치 한 것 은 ?',
'타이어 의 트레드 에 대한 설명 으로 틀린 것 은 ?',
'기중기 의 사용 용도 로 적합하지 않은 것 은 ?',
'화물 을 적재 하고 주행 할 때 포크 와 지면 과의 간격 으로 가장 적합한 것 은 ?',
'동력 전달 장치 에서 클러치 판 은 어떤 축의 스플라인 에 끼워져 있는가 ?',
'동력 전달 장치 에 사용 되는 차 동기 어 장치 에 대한 설명 으로 틀린 것 은 ?',
'타이어 식 로더 가 무한 궤도 식 로더 에 비해 가장 좋은 점 은 ?',
'유압식 굴삭기 의 주행 동력 으로 이용 되는 것 은 ?',
'건설 기계 의 소유자 는 다음 어느 령 이 정 하는 바 에 의하여 건설 기계 의 등록 을 하여야 하는가 ?',
'건설 기계 구 조 변경 범위 에 포함 되지 않는 사항 은 ?',
'도로 교통법 상 에서 교통 안전표지 의 구분 이 맞는 것 은 ?',
'도로교통법 상 서행 또는 일시 정지 할 장소 로 지정 된 곳 은 ?',
'원 동기 전문 건설 기계 정비 업 의 사업 범위 에 속 하지 않는 것 은 ?',
'긴급 자동차 의 우선 통행 에 관 한 설명 이 잘 못 된 것 은 ?',
'건설 기계 장비 의 제동장치 에 대한 정기 검사 를 면제 받고 자하 는 경우 첨부 하여야 하는 서류 는 ?',
'승차 인원 · 적재 중량 에 관 하여 안전 기준 을 넘어서 운행 하고자 하는 경우 누구 에게 허가 를 받아야 하는가 ?',
'건설 기계 의 조종 중 과실 로 100만원 의 재산 피해 를 입힌 때 면허 처분 기준 은 ?',

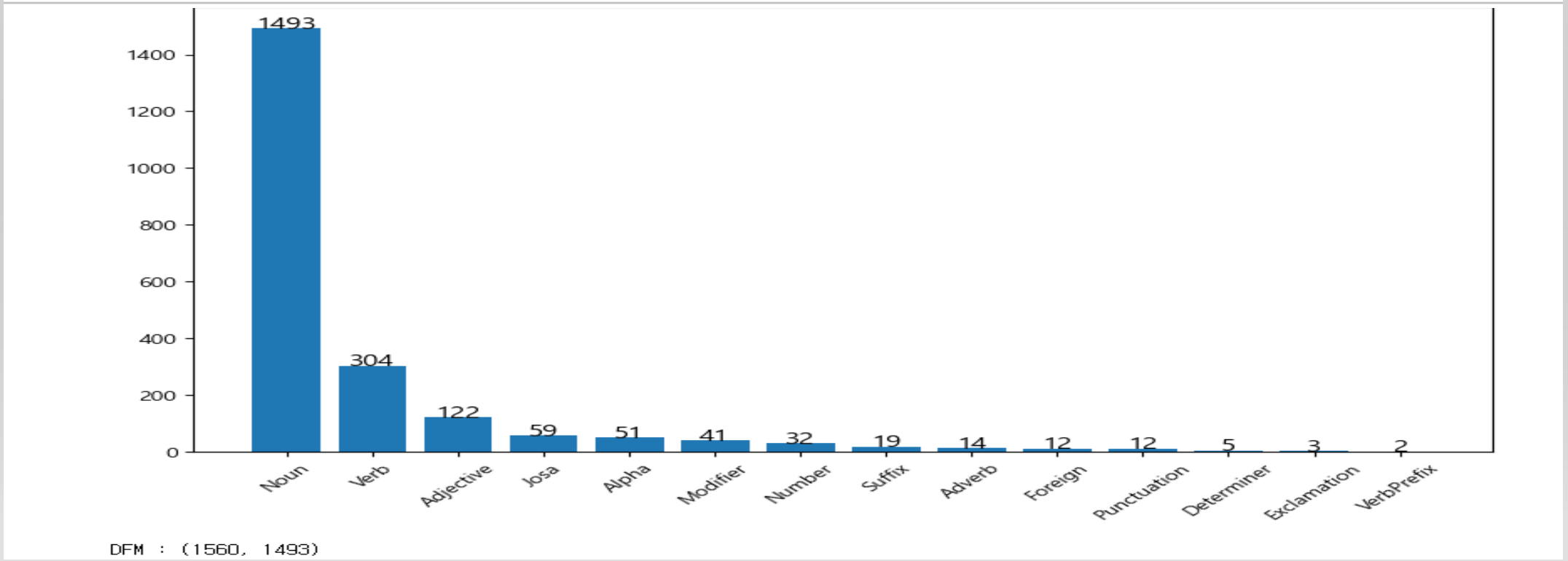
{('게', 'Noun'),
 ('경', 'Noun'),
 ('계', 'Noun'),
 ('공', 'Noun'),
 ('급', 'Noun'),
 ('덤프', 'Noun'),
 ('레드', 'Noun'),
 ('만', 'Noun'),
 ('매', 'Noun'),
 ('맥', 'Noun'),
 ('모', 'Noun'),
 ('반', 'Noun'),
 ('사', 'Noun'),
 ('석기', 'Noun'),
 ('성', 'Noun'),
 ('쇄', 'Noun'),
 ('순', 'Noun'),
 ('스플라', 'Noun'),
 ('식', 'Noun'),
 ('십', 'Noun'),
 ('액슬', 'Noun'),
 ('여', 'Noun'),
 ('연', 'Noun'),
 ('예', 'Noun'),
 ('오', 'Noun'),
 ('온', 'Noun'),
 ('요', 'Noun'),
 ('이나', 'Noun'),
 ('이지', 'Noun'),
 ('인', 'Noun'),
 ('작', 'Noun'),
 ('저', 'Noun'),
 ('조', 'Noun'),
 ('처럼', 'Noun'),
 ('첵', 'Noun'),
 ('하니', 'Noun')}

4. 그러면 형태소 기준으로 띄어쓰기를 하지 말고 원본 말뭉치 그대로 처리를 하는 방법은?
- TfidfVectorizer, CounterVectorizer에 tokenizer, vocabulary 파라미터를 사용하여 해결할 수 있다.

In [403]:
```
def my_tokenizer(doc) :
    return [
        token for token, pos in okt.pos(doc) if pos in ['Noun']
    ]
```

executed in 8ms, finished 03:25:09 2022-10-19

In [ ]:
```
TfidfVectorizer(tokenizer=my_tokenizer) # tokenizer 인자에 토큰 리스트를 반환하는 함수를 적으면 자동으로 토큰화를 해줌
```



DFM : (1560, 1493)

Okt가 문서의 품사를 Noun(명사)로 판별한 토큰들만 사전으로 등록되어 희소 벡터로 산출

영어로 된 사전도 넣고싶을 때 vocabulary 파라미터를 이용한다.

22.9[kV]배전선로에 근접하여 굴삭기 등 건설기계로 작업시 안전관리상 맞는 것은?
일반 도시가스사업자의 지하 배관을 설치 할 때 공동주택 등의 부지 내에서는 몇m 이상의 깊이에 배관을 설치해야 하는가?
겨울철에 사용하는 엔진오일은 여름철에 사용하는 엔진 오일보다 점도의 상태는 어떤 것이 좋은가?
디젤기관과 관계없는 것은?
건설기계장비 운전 시 계기판에서 냉각수량 경고등이 점등 되었다
4행정기관에서 엔진이 4000rpm일 때 분사펌프의 회전수는?

Ex)Kv, rpm으로 된 특이한 영어 단위와 한글 사이에 있는 영문은 어떻게 처리 할까?
파이썬 라이브러리 nltk.tokenize.RegexpTokenizer를 이용한다. 쉽게 설명하면 정규표현식으로 자연어 처리를 할 수 있는 클래스이다.

In [406]:
```
document = ['AC 발전기에서 전류가 발생되는 것은?']

for d in document :
    tokenizer_express = RegexpTokenizer("[a-zA-Z]+")
    token = tokenizer_express.tokenize(d.lower())

    print(token)
```

executed in 18ms, finished 03:50:28 2022-10-19

```
['ac']
```

1493개의 한국어 명사와 정규표현식으로 추출한 세 글자 이상의 영어 사전, 총 1526개로 벡터화 한 모습

```
new_text = ['라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?']

voca = set()

vectorize_sim_df = eng_vectorize_sim(data, okt, ['Noun'], TfidfVectorizer(tokenizer=my_tokenizer, vocabulary=voca), new_text)
vectorize_sim_df
```

executed in 47.7s, finished 03:54:16 2022-10-19

형태소 기준 태그 셋 : 1526
OrderedDict([('Noun', 1493)])

1493

DFM : (1560, 1526)

```
['actuator' 'alternator' 'arm' 'charger' 'comp' 'coupling' 'crank'
 'cylinder' 'driver' 'fire' 'flow' 'fluid' 'free' 'full' 'gpm' 'grader'
 'grow' 'ilo' 'key' 'lng' 'low' 'maintenance' 'mesh' 'miss' 'motor' 'mpa'
 'nox' 'plp' 'pressure' 'rpm' 'shaft' 'surge' 'turbo' '가공' '가까이' '가능' '가동'
 '가득' '가루' '가변' '가속' '가스' '가시거리' '가압' '가연성' '가열' '가이드' '가장' '가장자리' '가지'
 '가파른' '가해진' '각' '각각' '각도' '각종' '간' '간격' '간극' '감소' '감속' '감시' '감전' '감전사'
 '갑자기' '강관' '강산' '개' '개선' '개시' '개조' '개폐' '거나' '거래' '거리' '거부' '거의' '거품' '건'
 '건널목' '건설' '건축물' '걸때' '걸이' '검사' '검인' '것' '게이지' '겨울철' '격' '견인' '결과' '결부'
 '결정' '결함' '겹' '겹친' '경계' '경고' '경고표지' '경과' '경보기' '경사' '경사면' '경상' '경우' '경유'
 '경찰' '경찰관' '경찰서' '계기' '계단' '계속' '계약' '계통' '고' '고등' '고려' '고무' '고발' '고속도로'
 '고압' '고유' '고의' '고장' '고장원' '고정' '고착' '곡선' '골격' '곳' '공구' '공급' '공기' '공기청정기'
 '공동' '공동현' '공무원' '공사' '공유' '공장' '과' '과도' '과부' '과실' '과열' '과정' '과태료' '관'
 '관계' '관련' '관로' '관리' '관심' '관용' '관청' '관할' '광선' '교대' '교류' '교류발전기' '교부' '교육'
 '교차로' '교체' '교통' '교통법' '교통부' '교통사고' '교통안전표지' '교환' '구' '구급' '구내' '구동' '구로'
 '구류' '구별' '구분' '구비' '구성' '구성요소' '구의' '구인' '구입' '구조' '국부' '국제노동기구' '국토해양부'
 '굴삭' '굴삭기' '굴착' '권선' '권장' '궤도' '귀환' '규' '규정' '규제' '그' '그대로' '그리스' '그림'
 '근로' '근로자' '근접' '금' '금속' '금액' '금지' '급별' '급상승' '급속' '급유' '기' '기간' '기계'
 '기계로' '기관' '기구' '기기' '기능' '기동' '기동전' '기본' '기어' '기어오일' '기자' '기적' '기준'
 '기중기' '기타' '기한' '기호' '긴급' '길이' '깊이' '끝' '끼' '나' '나트륨' '낙하' '날' '날개' '날로'
 '납부' '납산' '낮' '내' '내부' '내용' '내의' '내일' '냄새' '냉각' '냉각수' '냉매' '너' '너비' '너트'
 '널리' '년' '노면' '노즐' '노출' '노크' '노킹' '녹색' '놀이기구' '농도' '높이' '누' '누구' '누산'
 '누설' '누유' '눈' '눈금' '는가' '능이' '다른' '다시' '다음' '다이오드' '다해' '단' '단동' '단위'
 '단자' '단점' '단지' '달' '당' '당사자' '대기' '대기압' '대비' '대상' '대용' '대작' '대책' '대체'
 '대피' '대하' '대한' '대해' '대행' '대형' '댐퍼' '더' '덤프트럭' '덮개' '도' '도관' '도로' '도로교통법'
 '도매' '도색' '도시' '도시가스' '도우' '도저' '도주' '도중' '도지사' '독촉' '돌' '돕기' '동' '동기'
 '동력' '동륜' '동부' '동시' '동식' '동안' '동작' '동파' '두' '뒤' '듀브' '드' '드라이버' '드라이브'
 '드래' '드럼' '드럼통' '드렌' '등' '등록' '등외' '등화' '디스크' '디젤' '디젤기관' '디젤엔진' '디콤프'
 '땅속' '때' '또' '라디에이터' '라면' '라이너' '라이닝' '라인' '란' '래시' '램프' '랩' '랭킹' '량' '런'
 '레버' '레이' '레이더' '렌치' '령' '로' '로더' '로드' '로부터' '로서' '로우' '로커' '로터' '로프'
```

이에 대한 유사도 평가를 높은 순으로 나열

	코사인 유사도	문서	유클리디안 거리	문서
303	0.836309	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?	0.572173	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?
7	0.836309	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?	0.572173	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?
6	0.692823	라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우의 원인은?	0.783807	라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우의 원인은?
1203	0.692823	라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우는 원인은?	0.783807	라디에이터 캡을 열었을 때 냉각수에 오일이 섞여있는 경우는 원인은?
1084	0.567157	기관이 작동 중 라디에이터 캡 쪽으로 물이 상승하면서 연소가스가 누출될 때의 원인에...	0.930422	기관이 작동 중 라디에이터 캡 쪽으로 물이 상승하면서 연소가스가 누출될 때의 원인에...
...	...	...	...	...
519	0.000000	다음에서 가장 높은 압력을 발생시키는 유압펌프의 형식은?	1.414214	다음에서 가장 높은 압력을 발생시키는 유압펌프의 형식은?
518	0.000000	호이스트형 유압호스 연결부분에 가장 많이 사용하는 것은?	1.414214	호이스트형 유압호스 연결부분에 가장 많이 사용하는 것은?
517	0.000000	유압탱크의 구비조건이 아닌 것은?	1.000000	유압탱크의 구비조건이 아닌 것은?
516	0.000000	밀폐된 용기 내의 일부에 가해진 압력은 어떻게 전달되는가?	1.414214	밀폐된 용기 내의 일부에 가해진 압력은 어떻게 전달되는가?
1559	0.000000	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?	1.000000	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?

1560 rows × 4 columns

명사, 동사, 형용사, 대한 유사도 평가를 높은 순으로 나열

Out[424]:

	코사인 유사도	문서	유클리디안 거리	문서
1263	0.595752	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?	0.899165	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
7	0.538066	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?	0.961181	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?
303	0.538066	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?	0.961181	압력식 라디에이터 캡에 대한 설명으로 적합한 것은?
1081	0.498708	디젤기관에서 실화할 때 나타나는 현상으로 옳은?	1.001292	디젤기관에서 실화할 때 나타나는 현상으로 옳은?
1455	0.466887	축전지 터미널에 부식이 발생하였을 때 나타나는 현상과 가장거리가 먼 것은?	1.032582	축전지 터미널에 부식이 발생하였을 때 나타나는 현상과 가장거리가 먼 것은?
...	...	...	...	...
521	0.000000	오일의 압력이 낮아지는 원인이 아닌 것은?	1.414214	오일의 압력이 낮아지는 원인이 아닌 것은?
520	0.000000	압력제어 밸브는 어느 위치에서 작동하는가?	1.000000	압력제어 밸브는 어느 위치에서 작동하는가?
519	0.000000	다음에서 가장 높은 압력을 발생시키는 유압펌프의 형식은?	1.414214	다음에서 가장 높은 압력을 발생시키는 유압펌프의 형식은?
518	0.000000	호이스트형 유압호스 연결부분에 가장 많이 사용하는 것은?	1.414214	호이스트형 유압호스 연결부분에 가장 많이 사용하는 것은?
1559	0.000000	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?	1.414214	굴착도중 전력케이블 표지시트가 나왔을 경우의 조치사항으로 적합한 것은?

1560 rows × 4 columns

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'

실험 결과

- tf-idf 벡터화 1-gram 일 때의 코사인 유사도, 유클리디안 거리 실험 결과

	코사인 유사도	문서
1086	0.636605	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
1263	0.636605	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
1242	0.579410	유압 계통에서 릴리프밸브 스프링의 장력이 약화 될 때 발생 될 수 있는 현상은?
525	0.532016	유압펌프의 압력조절밸브 스프링 장력이 높은 것을 사용하면 나타나는 현상으로 가장 적...
157	0.466839	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?

	유클리디안 거리	문서
1263	0.852520	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
1086	0.852520	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
1242	0.917159	유압 계통에서 릴리프밸브 스프링의 장력이 약화 될 때 발생 될 수 있는 현상은?
525	0.967454	유압펌프의 압력조절밸브 스프링 장력이 높은 것을 사용하면 나타나는 현상으로 가장 적...
157	1.032629	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'

실험 결과

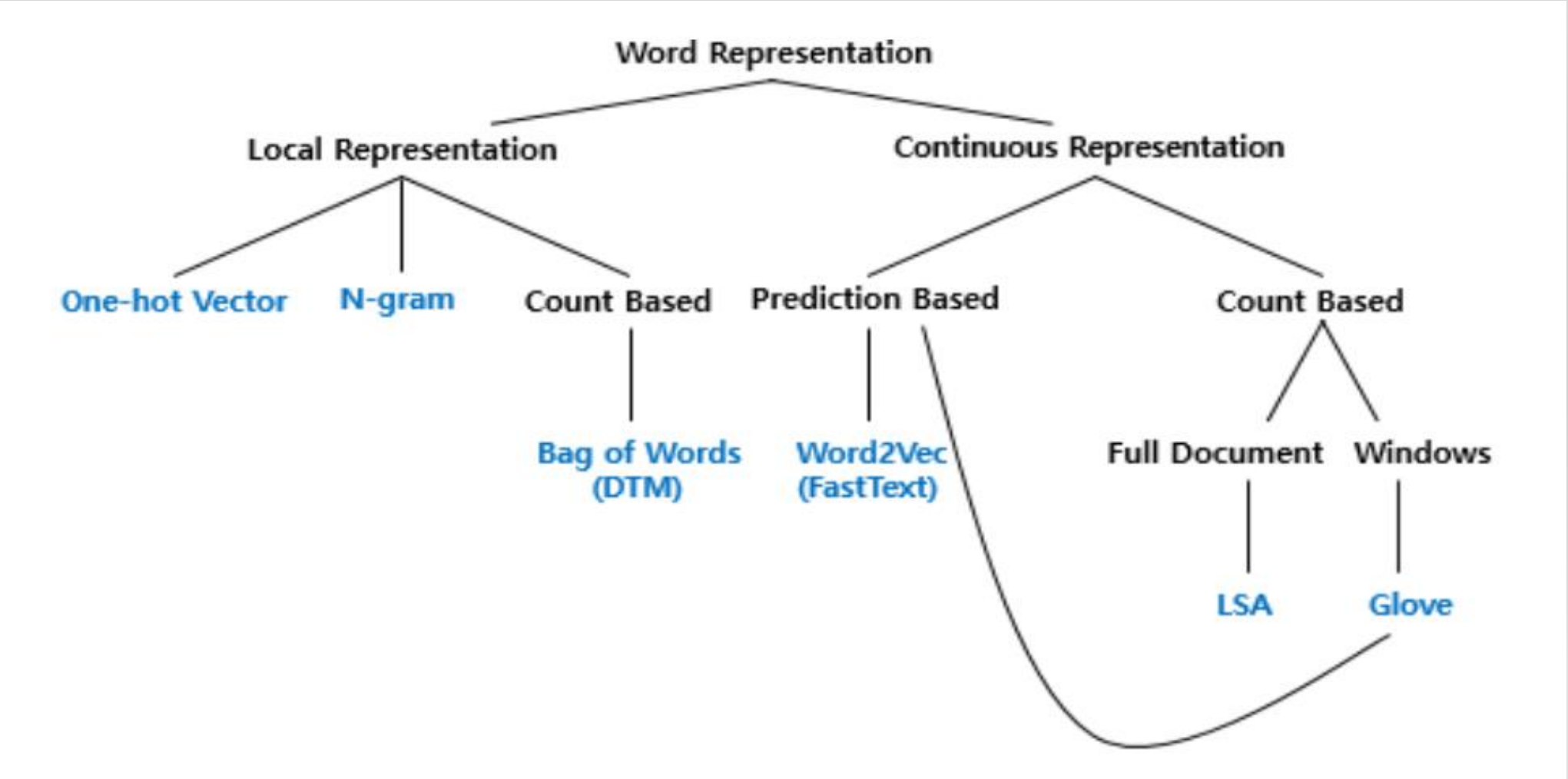
- tf-idf 벡터화 2-gram 일 때의 코사인 유사도, 유클리디안 거리 실험 결과

	코사인 유사도	문서
1242	0.429359	유압 계통에서 릴리프밸브 스프링의 장력이 약화 될 때 발생 될 수 있는 현상은?
157	0.328928	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
461	0.328928	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
525	0.281269	유압펌프의 압력조절밸브 스프링 장력이 높은 것을 사용하면 나타나는 현상으로 가장 적...
1285	0.207349	클러치 스프링의 장력이 약하면 일어날 수 있는 현상으로 가장 적합한 것은?

	유클리디안 거리	문서
510	1.000000	주.정차를 할 수 있는 곳은?
1242	1.068308	유압 계통에서 릴리프밸브 스프링의 장력이 약화 될 때 발생 될 수 있는 현상은?
157	1.158510	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
461	1.158510	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?
525	1.198942	유압펌프의 압력조절밸브 스프링 장력이 높은 것을 사용하면 나타나는 현상으로 가장 적...

정리

- 모델링 보다는 전처리에 집중하여 다양한 파라미터들을 적용해 봄
- 차후 LSA, RNN, Word2Vec을 진행하여 희소 벡터의 차원 축소와 시퀀스 기반의 모델을 익히는게 과제



Q & A