

입시용 유사 문제 찾기 프로젝트

프로젝트 수행 상황

- 9월 27일 ~ 현재 진행중
- NLP 용어 정리 하느라 3일 정도 날림
- 한국어 자연어 처리, 그 중에서 문장과 문장 사이의 유사도를 공부하는 중
- 발표 이후 연속 표현 기법 중 대표적인 word2vec와 자연어 처리 신경망(RNN , BERT) 공부 해야하는 상태 (솔직히 일주일 남아서 word2vec 개념 이해만 해도 성공적일듯.)

데이터는 어디서 구했니?

- 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

666 가장 자격증 기출문제 전자문제집 CBT - Chrome

comcbt.com/cbt/onlyview3.php

첫화면으로

전체 문제 보기 화면

지게차운전기능사 : 2011년10월09일 기출문제

1. 기관의 냉각팬에 대한 설명 중 틀린 것은?

1. 유체 커플링식은 냉각수의 온도에 따라서 작동된다.
2. 전동팬은 냉각수의 온도에 따라 작동된다.
3. 전동팬이 작동되지 않을 때는 물 펌프도 회전하지 않는다.
4. 전동팬의 작동과 관계없이 물 펌프는 항상 회전한다.

정답 : []

정답률 : 61%

<문제 해설>
전동팬과 물펌프는 별도로 각각 따로 움직입니다.
[해설작성자 : 나주공고 게토레이]

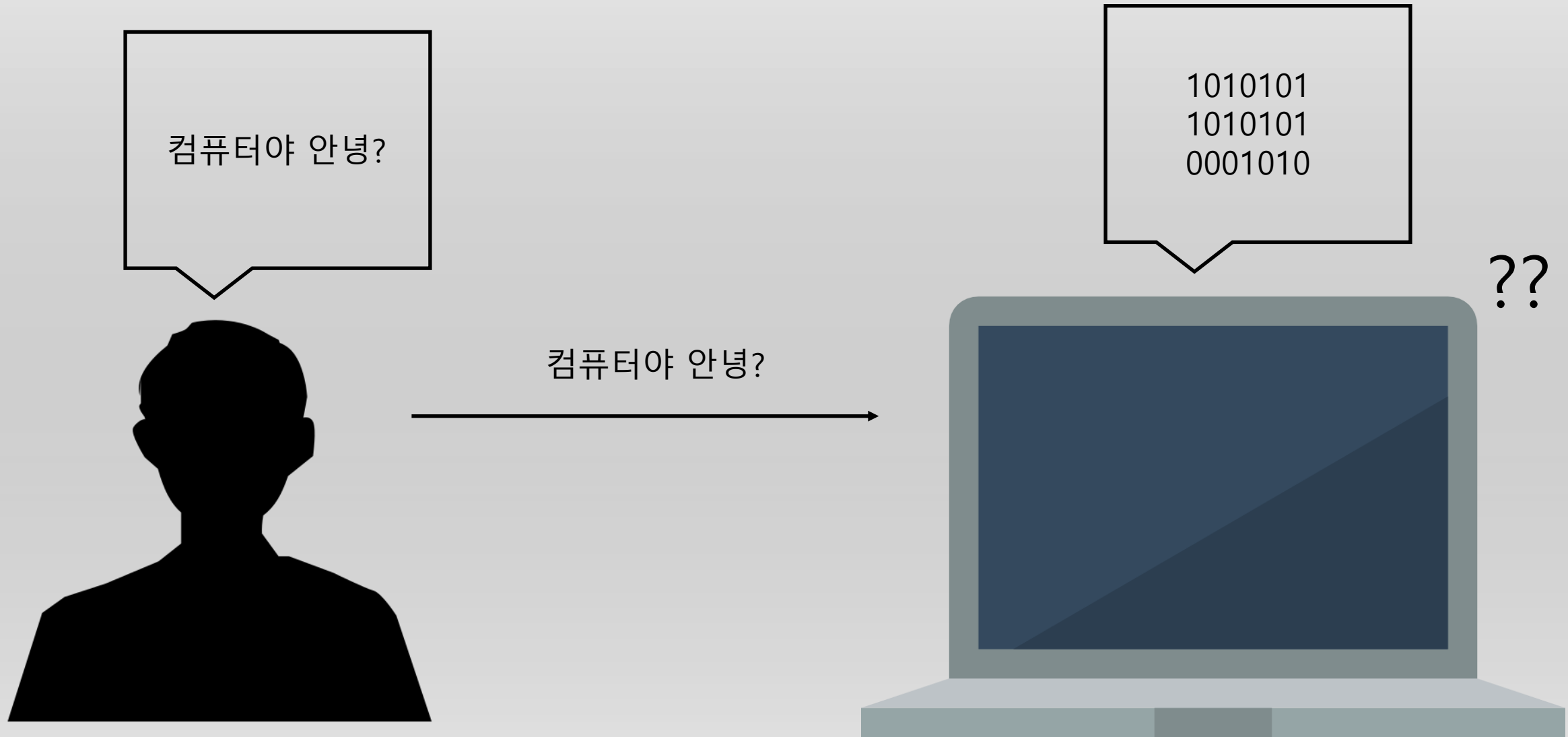
먼저, 기관의 냉각방식에는 전동팬과 냉각팬 방식의 두가지 방법이 있습니다. 전동팬은 모터로 구동을 하게 되는 것이고, 냉각팬은 기계식으로 팬벨트가 구동을 하게 됩니다. 냉각팬은 팬벨트에 연결되어 엔진이 구동되면 항상 회전을 하며, 냉각팬의 중심축은 물펌프 임펠러가 연결되어 냉각팬이 회전을 하면 물펌프도 같이 회전을 하게 됩니다. 전동팬은 모터를 사용하여 냉각수의 온도를 감지하여 온도가 높아졌을 때에만 작동을 하게 됩니다. 따라서 물펌프의 작동과 전동팬의 작동

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 적절한 것은?
윤활유 사용 방법으로 옳은 것은?
오일량은 정상이나 오일압력계의 압력이 규정치보다 높을 경우 조치사항 중 옳은 것은?
공기청정기의 설치 목적은?
발전기에서 발생하는 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 충전전 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 굴삭 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 며칠 이내에 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 펌핑시스템 등이 축류가 있으며 하루이 안력을 일정하게 유지시키는 밸브는?

비정형 데이터의 딜레마



1. 사용자가 원하는
문장을 배열 변환

컴퓨터야 안녕?

[1, 0, 1, 1]

2. 변환한 배열을 컴퓨터에게 전달

3. 사용자가
원하는 작업을
컴퓨터가 AI 모
델링 기반수행

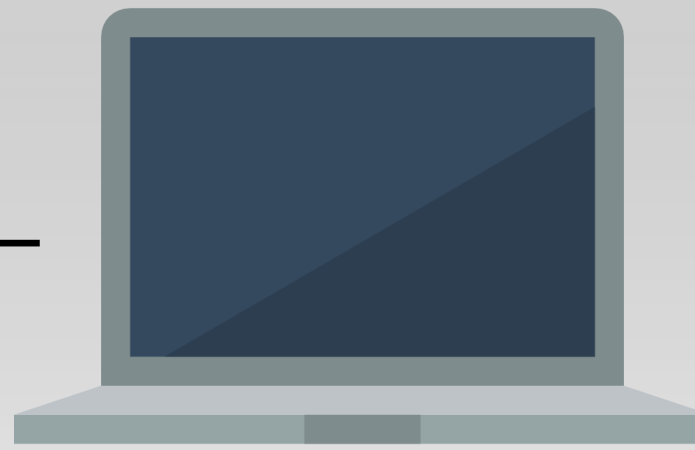
(해당 예는 질문
에 대한 대답을
하는 자연어 처
리 모델을 구현
했다고 가정)

4. 모델에 의해 출력 배열 [0, 1, 1, 0] 산출

[0, 1, 1, 0]

Conversion : '안녕!'

5. 출력 배열을 자연어로
처리하여 전달



텍스트 추출 및
클렌징



토큰화 및
N-gram



토큰화
피쳐 벡터라이징



모델링 기반
학습 예측
및 평가

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/cm²

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 적절한 것은?

윤활유 사용 방법으로 옳은 것은?

오일량은 정상이나 오일압력계의 압력이 규정치보다 높을 경우 조치사항 중 옳은 것은?

공기청정기의 설치 목적은?

발전기에서 발생하는 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 충전전 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 굴삭 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?

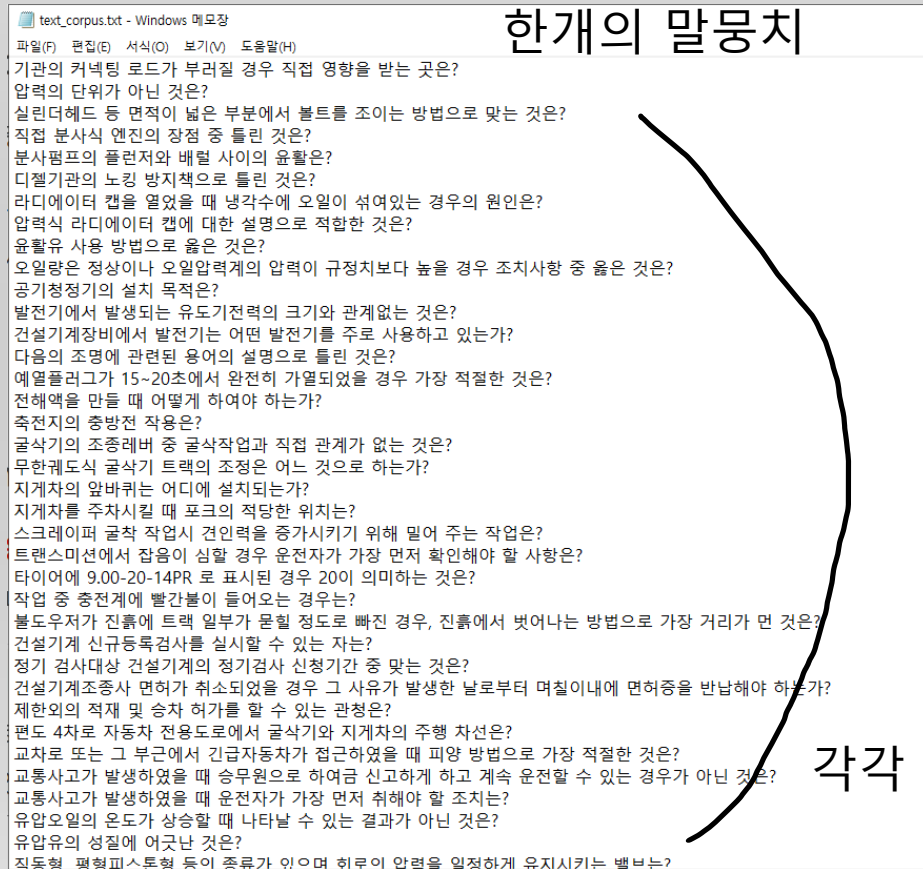
유압오일의 온도가 상승할 때 나타날 수 있는 결과가 아닌 것은?

유압유의 성질에 어긋난 것은?

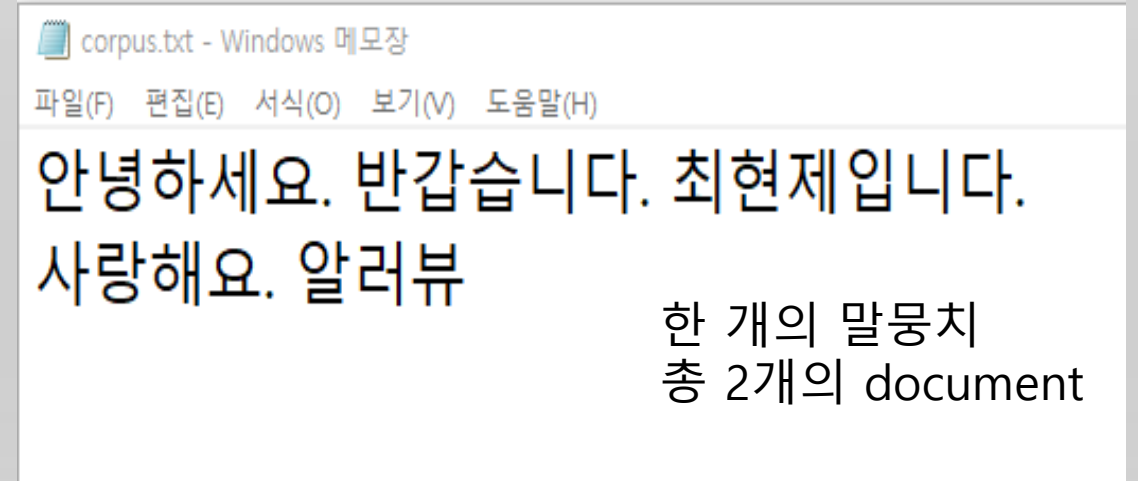
직동형 펌프피스톤형 등이 종류가 있으며 히로이 안력을 일정하게 유지시키는 밸브는?

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



각각 하나의 문장이 하나의 document

토큰화

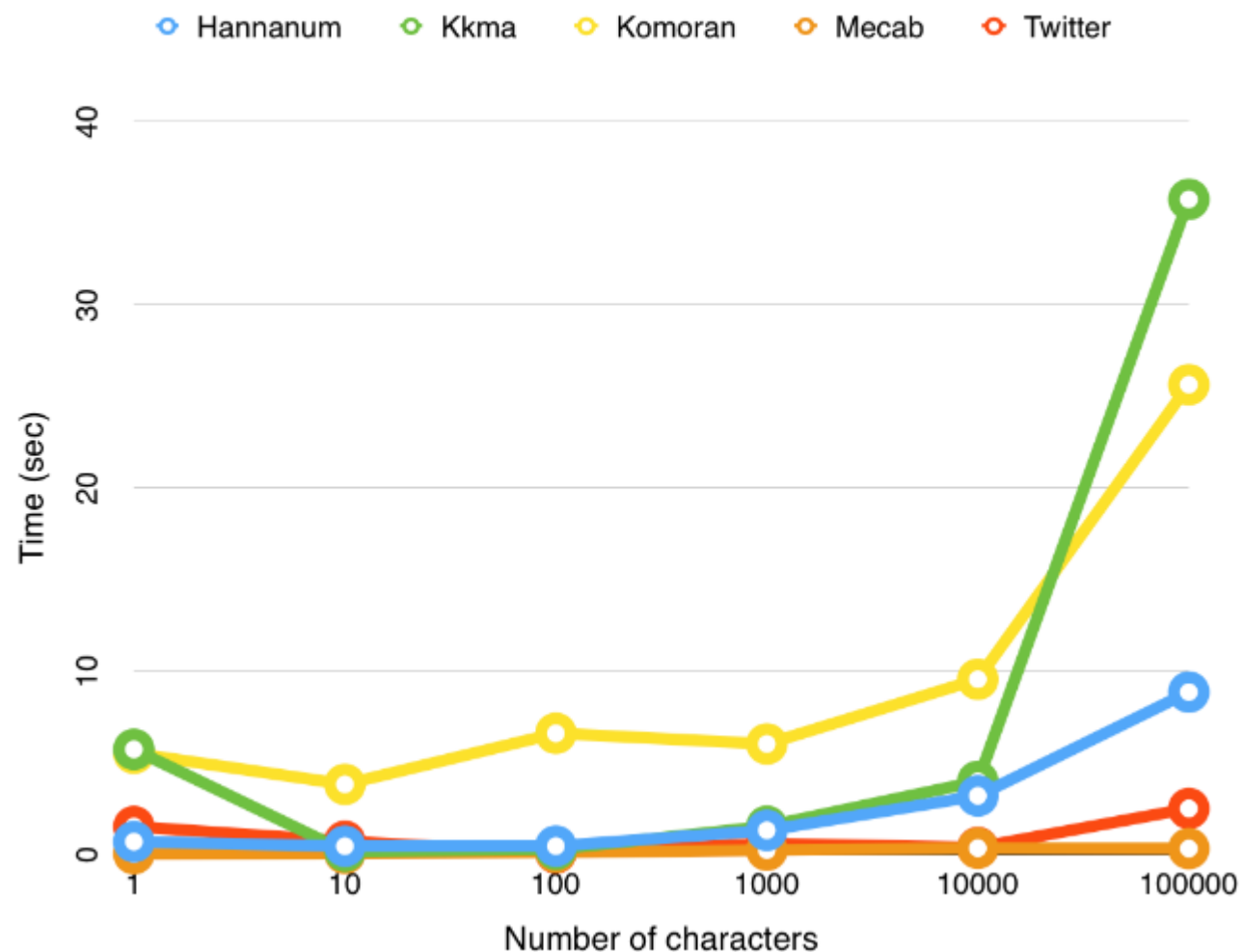
- 말뭉치(corpus)에서 문장마다 토큰이라 불리는 단위로 나누는 작업이다.
- 문장 토큰화, 단어 토큰화가 있다. 한국어는 단어 토큰화, 그 중에서 의미를 가지는 요소의 가장 작은 단위인 형태소 기준으로 토큰화를 한다.
- 파이썬에서는 konlpy(코엔엘파이) 모듈에서 5개의 형태소 분석기 중 하나를 쓴다.

```
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 : 57.23969388008118 sec
```

토큰화

문자의 개수를 늘려감에 따라 모든 클래스의 실행 시간은 기하급수적으로 증가합니다.



<https://konlpy-ko.readthedocs.io/ko/v0.4.3/morph/>

N-gram

- n개의 단어를 묶어 토큰화 한다.
- 기준 단어의 앞 단어와 뒤 단어가 하나의 어휘 일 때 성능이 향상, 어느 정도 순서가 보장 된다.
- 말뭉치의 토큰이 많을수록 피쳐 개수가 늘어난다.

document : '시베리아 호랑이가 어슬렁 거립니다.'

1-gram : '시베리아', '호랑이가', '어슬렁', '거립니다.'

2-gram : ('시베리아 호랑이가'), ('호랑이가' '어슬렁'), ('어슬렁 거립니다')

1-gram

```
tfidf_vector_fit = TfidfVectorizer(ngram_range=(1, 1)).fit(morphs_text_lst)
print(len(tfidf_vector_fit.vocabulary_))
# tfidf_vector = tfidf_vector.transform(morphs_text_lst)
# print(tfidf_vector.shape)
# print(tfidf_vector.vocabulary_)

# 20번 째 인덱스 -> 장점 : tf-idf 결과 20
# 22번 째 인덱스 -> 직접 : tf-idf 결과 22
```

executed in 157ms, finished 04:31:58 2022-10-12

1606

2-gram

```
tfidf_vector_fit = TfidfVectorizer(ngram_range=(2, 2)).fit(morphs_text_lst)
print(len(tfidf_vector_fit.vocabulary_))
# tfidf_vector = tfidf_vector.transform(morphs_text_lst)
# print(tfidf_vector.shape)
# print(tfidf_vector.vocabulary_)

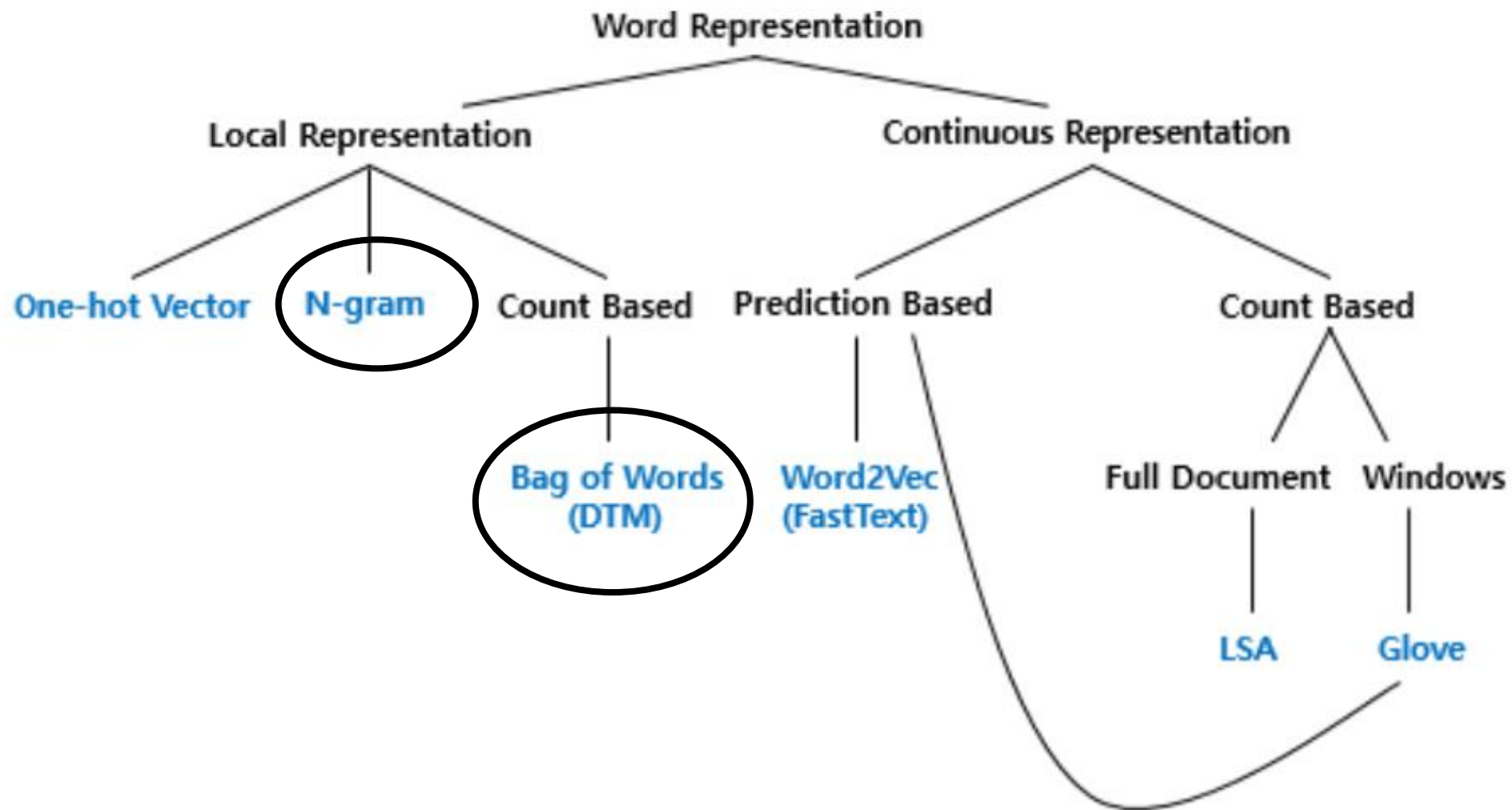
# 20번 째 인덱스 -> 장점 : tf-idf 결과 20
# 22번 째 인덱스 -> 직접 : tf-idf 결과 22
```

executed in 166ms, finished 04:32:32 2022-10-12

5175

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

IDF : $\ln\left(\frac{1+n}{1+df}\right) + 1$, 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

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

크기와 방향에 따른 유사도 측정

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

방향에 따른 유사도 측정

예시 [토끼 : 0], [거북이 : 1]
Document1 : 2000, 2000
Document2 : 2, 2

```
In [248]: euclidean_distances([[2, 2]], [[2000, 2000]])
```

```
executed in 12ms, finished 03:59:01 2022-10-12
```

```
Out [248]: array([[2825.59869762]])
```

```
In [250]: cosine_similarity([[2, 2]], [[2000, 2000]])
```

```
executed in 11ms, finished 03:59:40 2022-10-12
```

```
Out [250]: array([[1.]])
```

실험

1. n-gram

2. 벡터화

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'



건시시스템 <http://www.gunsys.com>

지게차운전기능사 필기 기출문제 (2013년 상시모의고사 3)

11. 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 |

'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...

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'

실험 결과

- CounterVectorize 벡터화일 때의 코사인 유사도, 유클리디안 거리 실험 결과

코사인 유사도		문서
1263	0.617213	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
1086	0.617213	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
525	0.524142	유압펌프의 압력조절밸브 스프링 장력이 높은 것을 사용하면 나타나는 현상으로 가장 적...
1242	0.478091	유압 계통에서 릴리프밸브 스프링의 장력이 약화 될 때 발생 될 수 있는 현상은?
157	0.455842	유압조정 밸브에서 조정 스프링의 장력이 클 때 나타나는 현상은?

유클리디안 거리		문서
1263	2.236068	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나는 현상은?
1086	2.236068	라디에이터 캡의 스프링이 파손 되었을 때 가장 먼저 나타나 는 현상은?
1380	2.828427	압력의 단위가 아닌 것은?
510	2.828427	주.정차를 할 수 있는 곳은?
1	2.828427	압력의 단위가 아닌 것은?

'라디에이터 캡의 압력스프링 장력이 약화되었을 때 나타나는 현상은?'

실험 결과

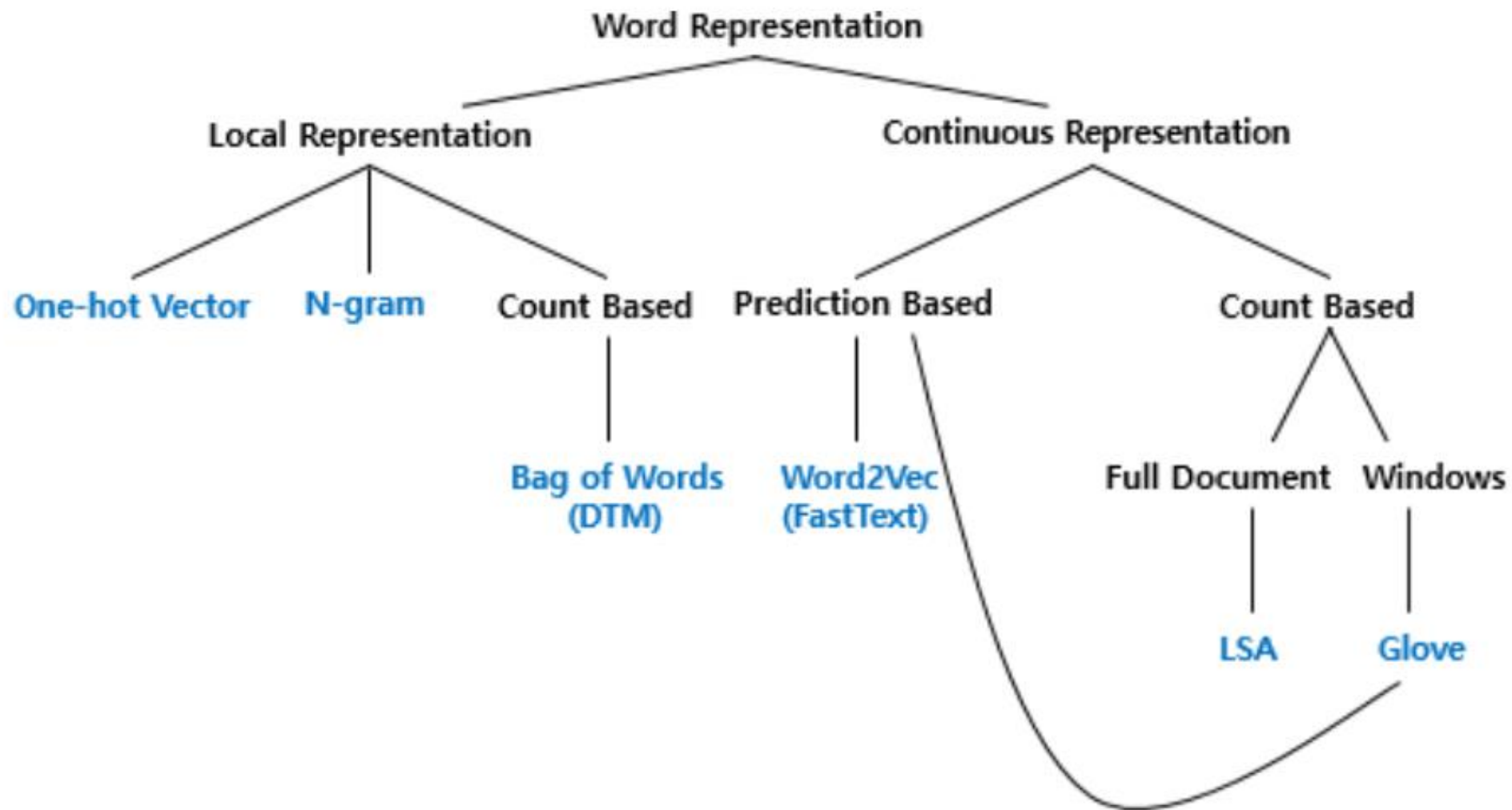
- tf-idf 벡터화일 때의 코사인 유사도, 유클리디안 거리 실험 결과

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?

사실은... 희소 행렬의 일부분을 쓴 것 입니다...

일주일 안에 word2Vec과 자연어 관련 인공지능망 모델 공부해야 하는데... 최대한 해볼게요.



Q & A