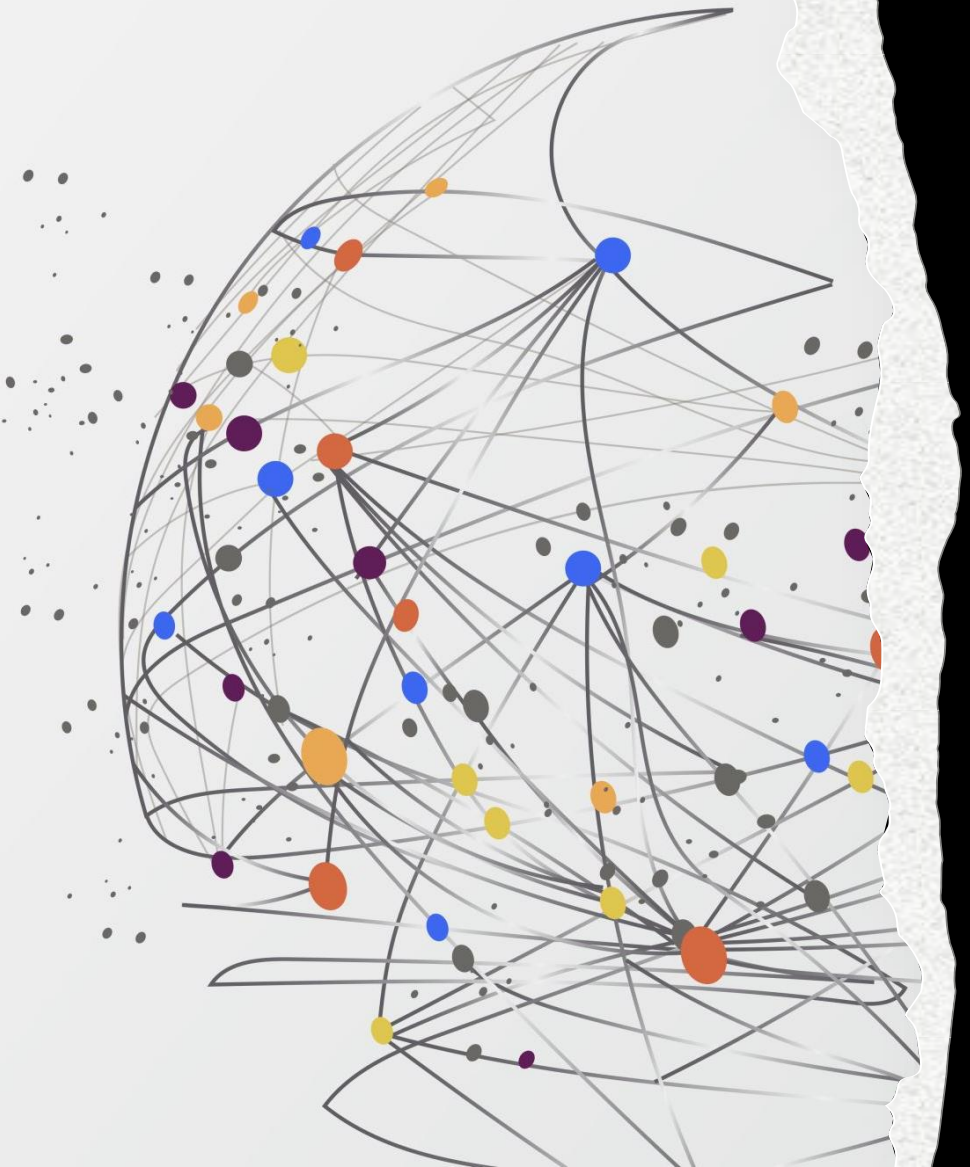


7. 임베딩(EmbeddingEmbedding)

김석현 21016006

Embedding?

:사람이 쓰는 자연어를 기계가 이해할 수 있는 숫자의 나열인 벡터로 바꾼 결과 혹은 그 과정 전체를 의미

단어-문서 행렬

구분	문서 A	문서 B	문서 C
단어 i	0	3	5
단어 j	1	0	0
단어 k	1	0	0

- A의 임베딩은 [0,1,1]
- 단어 i의 임베딩은 [0,3,5]

Embedding method

- 인코딩

-정수 인코딩

-원 핫 인코딩

- Word2Vec

정수 인코딩

text="평생 살 것처럼 꿈을 꾸어라. 그리고 내일 죽을 것처럼 오늘을 살아라."

정수 인코딩

2 3 1 4 5 6 7 8 1 9 10

"평생 살 것처럼 꿈을 꾸어라. 그리고 내일 죽을 것처럼 오늘을 살아라."

정수 인코딩

{'것처럼': 1, '평생': 2, '살': 3, '꿈을': 4, '꾸어라': 5, '그리고': 6, '내일': 7, '죽을': 8, '오늘을': 9, '살아라': 10}

원 핫 인코딩

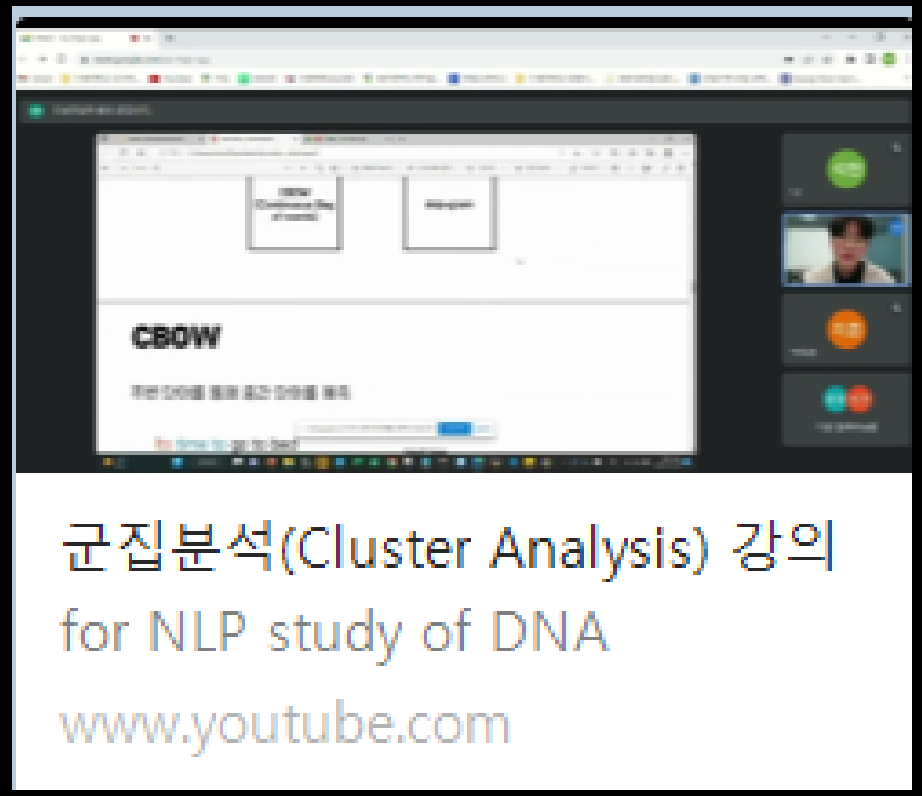
You say goodbye and I say hello

단어	단어 인덱스	원-핫 벡터
you	0	[1, 0, 0, 0, 0, 0, 0]
say	1	[0, 1, 0, 0, 0, 0, 0]
goodbye	2	[0, 0, 1, 0, 0, 0, 0]
and	3	[0, 0, 0, 1, 0, 0, 0]
I	4	[0, 0, 0, 0, 1, 0, 0]
say	5	[0, 0, 0, 0, 0, 1, 0]
hello	6	[0, 0, 0, 0, 0, 0, 1]

원 핫 인코딩

Ex) 강아지 **= [0 0 0 0 1 0 0 0 0 0 0 0 ... 중략 ... 0]**

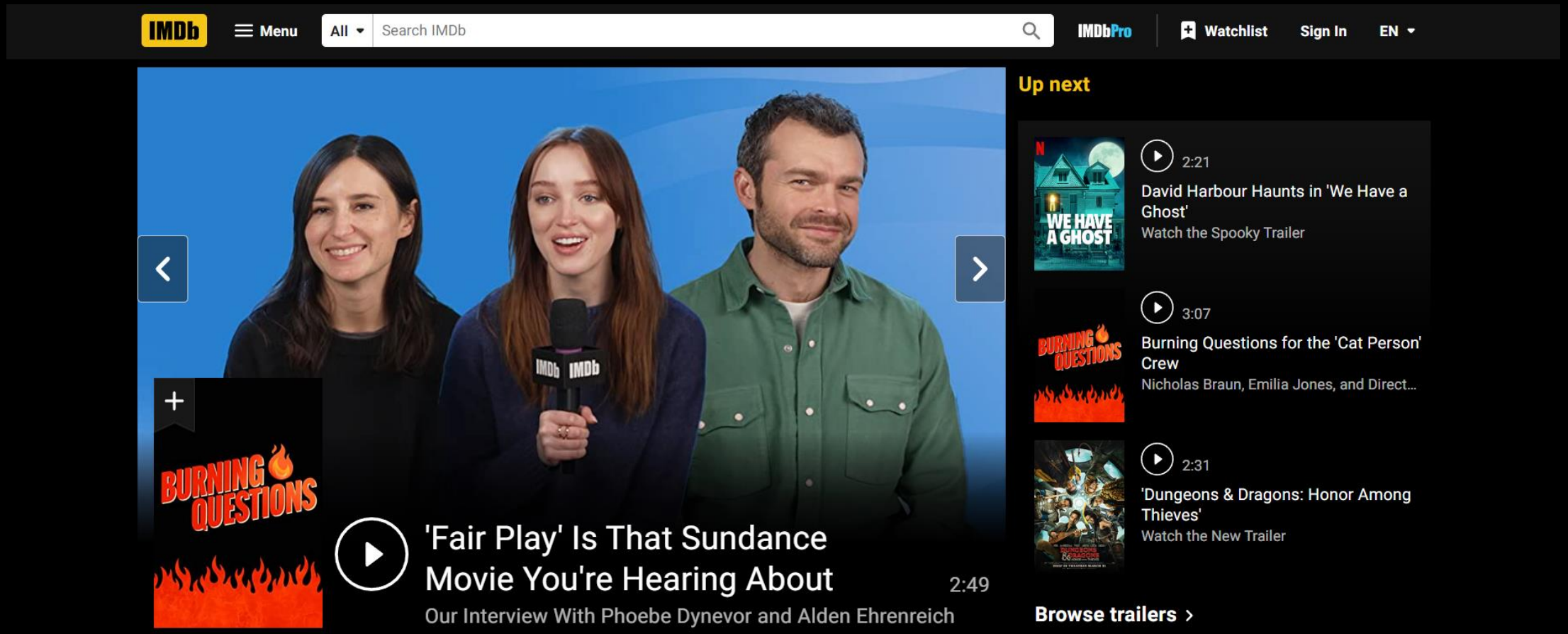
Word2Vec



https://www.youtube.com/watch?v=E7sBbILgSM4

IMDB

Internet Movie Database



```
=imdb.load_data(num_words=num_words)
```

```
array([list([1, 14, 22, 16, 43, 530, 973, 2, 2, 65, 458, 2, 66, 2, 4, 173, 36, 256, 5, 25, 100, 43, 838, 112, 50, 670, 2, 9, 35, 480, 284, 5, 150, 4, 172,
112, 167, 2, 336, 385, 39, 4, 172, 2, 2, 17, 546, 38, 13, 447, 4, 192, 50, 16, 6, 147, 2, 19, 14, 22, 4, 2, 2, 469, 4, 22, 71, 87, 12, 16, 43, 530, 38,
76, 15, 13, 2, 4, 22, 17, 515, 17, 12, 16, 626, 18, 2, 5, 62, 386, 12, 8, 316, 8, 106, 5, 4, 2, 2, 16, 480, 66, 2, 33, 4, 130, 12, 16, 38, 619, 5, 25,
124, 51, 36, 135, 48, 25, 2, 33, 6, 22, 12, 215, 28, 77, 52, 5, 14, 407, 16, 82, 2, 8, 4, 107, 117, 2, 15, 256, 4, 2, 7, 2, 5, 723, 36, 71, 43, 530, 476,
26, 400, 317, 46, 7, 4, 2, 2, 13, 104, 88, 4, 381, 15, 297, 98, 32, 2, 56, 26, 141, 6, 194, 2, 18, 4, 226, 22, 21, 134, 476, 26, 480, 5, 144, 30, 2, 18,
51, 36, 28, 224, 92, 25, 104, 4, 226, 65, 16, 38, 2, 88, 12, 16, 283, 5, 16, 2, 113, 103, 32, 15, 16, 2, 19, 178, 32]),
       list([1, 194, 2, 194, 2, 78, 228, 5, 6, 2, 2, 2, 134, 26, 4, 715, 8, 118, 2, 14, 394, 20, 13, 119, 954, 189, 102, 5, 207, 110, 2, 21, 14, 69, 188,
8, 30, 23, 7, 4, 249, 126, 93, 4, 114, 9, 2, 2, 5, 647, 4, 116, 9, 35, 2, 4, 229, 9, 340, 2, 4, 118, 9, 4, 130, 2, 19, 4, 2, 5, 89, 29, 952, 46, 37, 4,
455, 9, 45, 43, 38, 2, 2, 398, 4, 2, 26, 2, 5, 163, 11, 2, 2, 4, 2, 9, 194, 775, 7, 2, 2, 349, 2, 148, 605, 2, 2, 15, 123, 125, 68, 2, 2, 15, 349, 165, 2,
98, 5, 4, 228, 9, 43, 2, 2, 15, 299, 120, 5, 120, 174, 11, 220, 175, 136, 50, 9, 2, 228, 2, 5, 2, 656, 245, 2, 5, 4, 2, 131, 152, 491, 18, 2, 32, 2, 2,
14, 9, 6, 371, 78, 22, 625, 64, 2, 9, 8, 168, 145, 23, 4, 2, 15, 16, 4, 2, 5, 28, 6, 52, 154, 462, 33, 89, 78, 285, 16, 145, 95]),
       list([1, 14, 47, 8, 30, 31, 7, 4, 249, 108, 7, 4, 2, 54, 61, 369, 13, 71, 149, 14, 22, 112, 4, 2, 311, 12, 16, 2, 33, 75, 43, 2, 296, 4, 86, 320,
35, 534, 19, 263, 2, 2, 4, 2, 33, 89, 78, 12, 66, 16, 4, 360, 7, 4, 58, 316, 334, 11, 4, 2, 43, 645, 662, 8, 257, 85, 2, 42, 2, 2, 83, 68, 2, 15, 36, 165,
2, 278, 36, 69, 2, 780, 8, 106, 14, 2, 2, 18, 6, 22, 12, 215, 28, 610, 40, 6, 87, 326, 23, 2, 21, 23, 22, 12, 272, 40, 57, 31, 11, 4, 22, 47, 6, 2, 51, 9,
170, 23, 595, 116, 595, 2, 13, 191, 79, 638, 89, 2, 14, 9, 8, 106, 607, 624, 35, 534, 6, 227, 7, 129, 113]),
```

```
word_index = imdb.get_word_index()
reverse_word_index = dict(
    [(value, key) for (key,value) in word_index.items()]
)
decoded_review = ' '.join(
    [reverse_word_index.get(i-3, "?") for i in x_train[0]] # 0,1,2는 '패딩', '문서 시작', '사전에 없음'을 위한 인덱스이므로 3을 뺌
)
print(decoded_review)
```

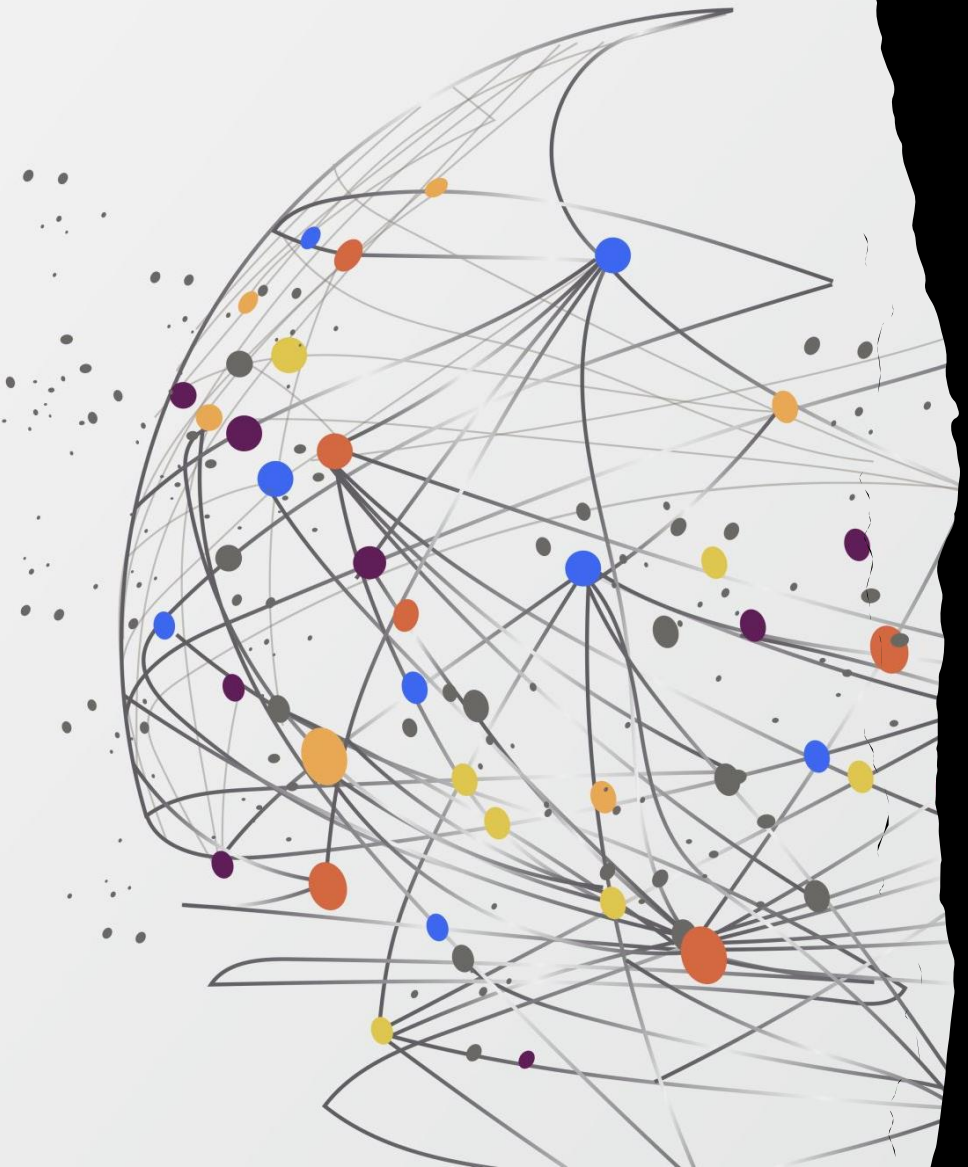
? this film was just brilliant casting ? ? story direction ? really ? the part they played and you could just imagine being there robert ? is an amazing actor and now the same being director ? father came from the same ? ? as myself so i loved the fact there was a real ? with this film the ? ? throughout the film were great it was just brilliant so much that i ? the film as soon as it was released for ? and would recommend it to everyone to watch and the ? ? was amazing really ? at the end it was so sad and you know what they say if you ? at a film it must have been good and this definitely was also ? to the two little ? that played the ? of ? and paul they were just brilliant children are often left out of the ? ? i think because the stars that play them all ? up are such a big ? for the whole film but these children are amazing and should be ? for what they have done don't you think the whole story was so ? because it was true and was ? life after all that was ? with us all

T-SNE

- https://gaussian37.github.io/ml-concept-t_sne/#

과제

- https://dacon.io/en/codeshare/4642



감사합니다.