T.C.

Firat University

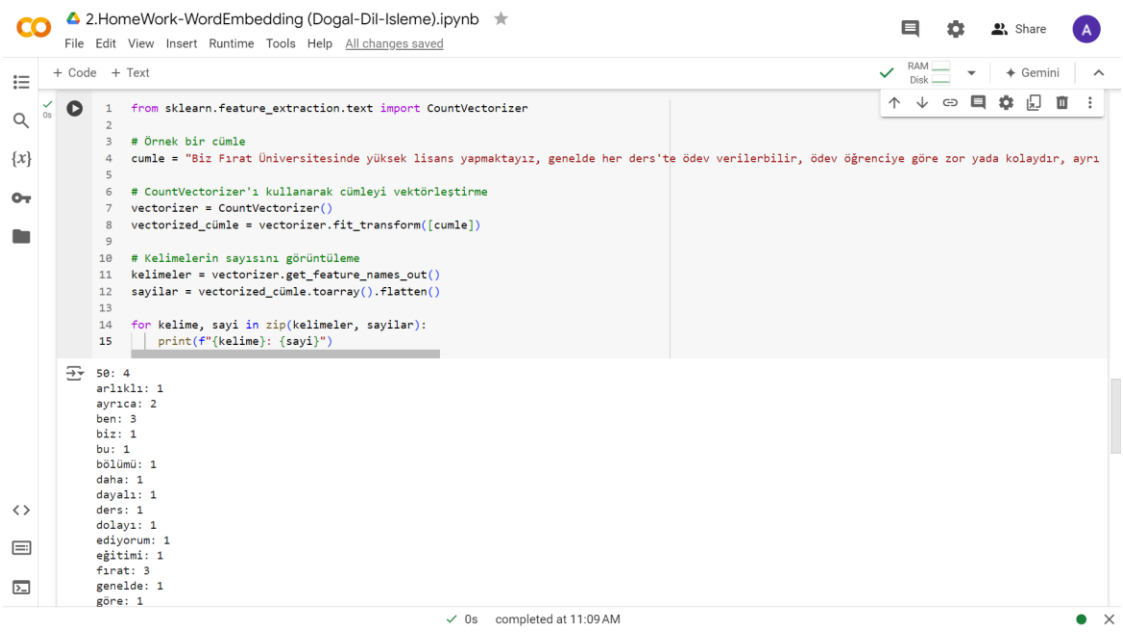
Mohammad Amin ASLAMI - 232137101

2. Homework - Word Embedding

Lesson: Natural Language Processing

Supervisor: Ass.Prof.Doc Murat AYDOĞAN

Cümlede geçen sayı ve kelimeleri, kaç defa geçtini öğrenme ve yazdırma



Şekil 1.1 Cümlede geçen sayı ve kelime, kaç defa geçtini öğrenme

Yukardaki kod da, ilk önce sayıları yazılır, ondan sonra kelime yazılır.

Örnek:

50: Sayısı 4 defa cümle içersinde geçmiştir.

ayrıca: 2 defa geçmiştir, **fırat:** 3 defa geçmiştir.

GloVe: Global Vectors for Word Representation

"**glove.6B.50d.txt**" google colab indirmek için aşağıdaki komutları kullandım.

1. !wget http://nlp.stanford.edu/data/glove.6B.zip
2. !unzip glove.6B.zip

Sonra aşağıdaki kodu çalıştırıp, düzgün bir şekilde çalıştı.

```
[37]  1  from gensim.models import KeyedVectors
      2
      3  # GloVe'nin önceden eğitilmiş modelini yükleme
      4  glove_file = 'glove.6B.50d.txt'
      5  glove_model = KeyedVectors.load_word2vec_format(glove_file, binary=False, no_header=True)
```

Şekil 2.1 Glove.6B Indirimi

Benzer kelimeleri çıkartma

Burada bir kelimenin eşanlamlarını çıkartmaya çalışıyoruz.

```
1  # Burada (book) esenin yada benzer kelimleri çıkartacaktır.
2  similar_words_glove = glove_model.most_similar('book', topn=10)
3  print("Benzer kelimeler:")
4  for word in similar_words_glove:
5      print(word)
```

```
Benzer kelimeler:
('books', 0.9047631025314331)
('story', 0.8662747144699097)
('novel', 0.8550738096237183)
('writing', 0.843974232673645)
('published', 0.8439115881919861)
('biography', 0.8398316502571106)
('author', 0.8371229767799377)
('wrote', 0.8293616771697998)
('written', 0.8216683864593506)
('titled', 0.8155251145362854)
```

Şekil 3.1 Mesela (book) benzer kelimlerin çıkartma

Benzerlik Skoru

Queen sonucu skoru: **0.8523604273796082**

```
result = glove_model.most_similar(positive=['king', 'woman'], negative=['man'], topn=5)

# Sonucu yazdırma
print(f"'king - man + woman' işleminin sonucu: {result[0][0]} (Benzerlik Skoru: {result[0][1]})")
```

```
'king - man + woman' işleminin sonucu: queen (Benzerlik Skoru: 0.8523604273796082)
```

Şekil 3.2 Benzerlik skoru (queen)

Benzerlik Skoru

King sonucu skoru: **0.8523604273796082**

King çok az fark ile daha iyi bir sonuç vermiştir.

```
result = glove_model.most_similar(positive=['queen', 'man'], negative=['woman'], topn=5)

# Sonucu yazdırma
print(f"'queen - woman + man' işleminin sonucu: {result[0][0]} (Benzerlik Skoru: {result[0][1]})")
```

```
'queen - woman + man' işleminin sonucu: king (Benzerlik Skoru: 0.8612024784088135)
```

Şekil 3.3 Benzerlik skoru (queen)

Software Vektoru

```
software_vektoru = model.wv['software']
print(software_vektoru)
```

```
[ 0.00022947  0.00614663 -0.01365961 -0.00276419  0.01534664  0.01469861
 -0.00731165  0.00533572 -0.01665603  0.01239867 -0.00927548 -0.0064176
  0.01866365  0.00179289  0.01493189 -0.01212172  0.0103893   0.01984986
 -0.01695286 -0.01031119 -0.01410299 -0.00970072 -0.00753695 -0.01704524
  0.0159442  -0.00964603  0.0168673   0.01054826 -0.0131234   0.00793291
  0.01097174 -0.01488653 -0.01481117 -0.0049783  -0.01730343 -0.00317906
 -0.00080648  0.00660388  0.00292227 -0.00175649 -0.01116442  0.00344443
 -0.00177557  0.01357014  0.00798727  0.00905791  0.00288245 -0.00545181
 -0.00868396 -0.00205627]
```

Şekil 3.4 Software vektoru

Project codes: https://colab.research.google.com/drive/1mLUf2uG_jRp_DozxSUWX2lFcPHcvkJPp#scrollTo=hRpyVMz2e4kZ

Dataset: https://huggingface.co/datasets/maydogan/TRSAv1

Glove Embedding in Colab: https://www.youtube.com/watch?v=HniyqWFegTg

Edit by: Amin ASLAMI