

|  |  |
| --- | --- |
| DOĞAL DİL İŞLEME DÖNEM ÖDEVİ | |
| **VERİ SETİ ZENGİNLEŞTİRME** |  |
| MURAT DÖĞÜM **DERS**:  **DOĞAL DİL İŞLEME**  **DR. ÖĞR. ÜYESİ EDİP SERDAR GÜNER** |  |

İçindekiler

GOOGLE TRANSLATE KULLANARAK VERİ SETİNİN ÇOĞALTILMASI **3**

Kullanacağımız kütüphaneler **3**

Kod anlatımı **3**

Örnek dataset **5**

Kod **7**

Başvurular **7**

**GOOGLE TRANSLATE KULLANARAK VERİ SETİNİN ÇOĞALTILMASI**

Metin üretme çalışmaları, mevcut cümlelerin analiz edilerek yeni bilgi çıkarımlarının yapıldığı, var olan bir bilgiden yola çıkarak bununla bağlantılı anlamlı bilgilerin elde edildiği sistemlerdir. Bu sistemler, özellikle internet üzerinden yapılan aramalarda girilen cümlelerin türetilerek kullanıcılara arama sonuçları ile ilgili daha anlamlı sonuçlar döndürebilmek gibi kolaylıklar sağlarlar. Bizde Google ‘ın translate apisini kullanarak oluşturduğumuz data seti kullanarak veri çoğaltımı yapacağız ve bunu ikinci bir data set olarak kaydedeceğiz.

**Kullanacağımız kütüphaneler:**

1. googletrans
2. openpyxl

İlk önce Google translate kütüphanesini entegre edelim.

*pip install googletrans==4.0.0rc1*

Excel kütüphanesinde yükleyelim.

*pip install openpyxl*

Daha sonra import edelim,

*from googletrans import Translator*

Fonksiyonu çağıralım,

translator = Translator()

Cümleyi verelim

tr=" sonra gerekli cümleleri verip her dönütte tekrar edene kadar hepsini kaydedecek kod bittiğinde elinizde yeni bir data kümesi olacak "

Cümleyi while döngüne alıp tekrar edene kadar çıktı alalım

while(i<5):

txt = translator.translate(tr, src='tr', dest='en').text

tr = translator.translate(txt, src='en', dest='tr').text

if(tr==tr2):

break

elif(tr==cumle):

break

else:

ws2.append([tr,konu])

print(str(i)+":"+txt)

print(str(i)+":"+tr)

tr2=tr

i=i+1

Kod bittiğinde aşığımdaki gibi gözükecektir

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Çıktı aşağıdaki gibi olacaktır

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şimdi bu işlemi fonksiyon olarak kaydedelim. Her cümle için bu fonksiyonu kullanarak işlem yapacağız.

Data.xlsx adında bir veri kümesi oluşturduk ve excel kütüphanesi import ettik. Cümle ve konu olarak iki sütun olarak aldık sırayla bu cümleleri oluşturduğumuz fonksiyonla tek tek çoğaltıp oluşturduğumuz ikinci veri setinin(data2.xlsx) içine yazalım.

from openpyxl import Workbook,load\_workbook

wb = load\_workbook("data.xlsx")

ws = wb.active

wb2 = Workbook()

ws2 = wb2.active

Excelde veri okuma işlemi için aşağıdaki kodu kullandık oluşturduğumuz fonksiyonun içine sırayla cümleleri gönderdik

i=1

for x in ws:

cevir(ws[("A"+str(i))].value,ws[("B"+str(i))].value)

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

i=i+1

Excele yeni cümle eklemek için aşağıdaki kodu kullanalım

ws2.append([cumle,konu])

En son kaydedelim

wb2.save("data2.xlsx")

**ÖRNEK DATA.XLSX:**

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**ÇIKTI DATA2.XLSX:**

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**KONSOL ÇIKTISI:**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

KODUN SON HALI:

from googletrans import Translator

from openpyxl import Workbook,load\_workbook

wb = load\_workbook("data.xlsx")

ws = wb.active

wb2 = Workbook()

ws2 = wb2.active

def cevir(cumle,konu):

translator = Translator()

tr=cumle

print(konu)

ws2.append([cumle,konu])

tr2=""

i=1

while(i<5):

txt = translator.translate(tr, src='tr', dest='en').text

tr = translator.translate(txt, src='en', dest='tr').text

if(tr==tr2):

break

elif(tr==cumle):

break

else:

ws2.append([tr,konu])

print(str(i)+":"+txt)

print(str(i)+":"+tr)

tr2=tr

i=i+1

i=1

for x in ws:

cevir(ws[("A"+str(i))].value,ws[("B"+str(i))].value)

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

i=i+1

wb2.save("data2.xlsx")

# Başvurular

Blog, K. (tarih yok). *Python İle Excel Dosya (XLSX ve CSV) Satır Satır Okuma ve Yazma İşlemleri*. Kerteriz Blog: https://kerteriz.net/python-ile-excel-dosya-islemleri/ adresinden alındı