



BİTİRME PROJESİ

Haftalık Rapor – 29.04.2022

29 NİSAN 2022

KIRIKKALE ÜNİVERSİTESİ – BİLGİSAYAR MÜHENDİSLİĞİ İÖ

AHMET MUNGAN – 160255081

İÇİNDEKİLER

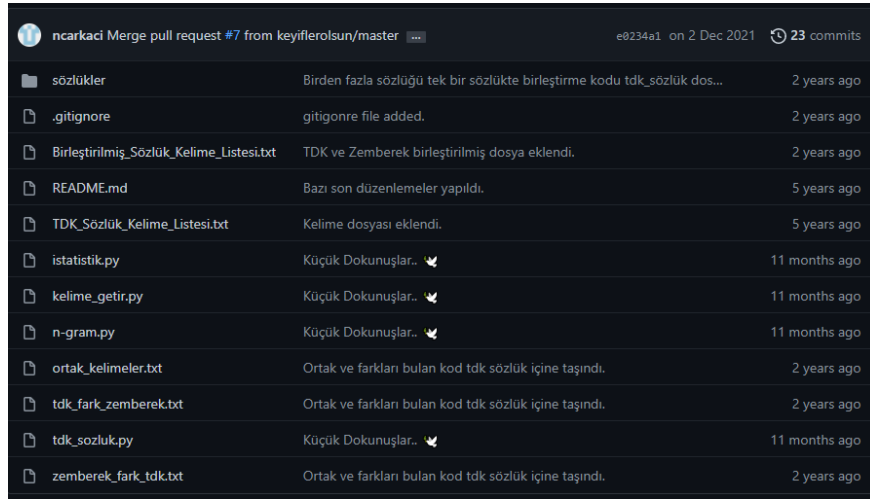
ÖZET.....	2
SÖZLÜK OLUŞTURMA İŞLEMLERİ	3
REFERANS VE KAYNAKÇA	5
EKLER.....	6

ÖZET

TDK'dan alınan verilerin işlenerek iyi hale getirildiği bir sözlük olarak text dosyası elde edilmiştir. Bu sözlüğün parçalara bölünmesi ve bu işlemlerin nedeni belirtilmiştir. Projenin geleceği için ne gibi avantajlar veya dezavantajlar barındırdığı yorumlanmıştır.

SÖZLÜK OLUŞTURMA İŞLEMLERİ

Türkçe diline yıllardan beridir ışık tutmuş TDK, yayınladığı sözlük ile doğruluğu temsil etmektedir. Hazır kütüphanelerin kullanılmadığı senaryolarda, köken belirlemekte bu kelimeler önem arz etmektedir. Dolayısıyla TDK'dan veri toplama işlemlerine girilebilir. Fakat bu işlemi daha evvel yapan, başarılı sonuçlar alan bir proje yaratıcısının github repository'sinde bulmak mümkün.



The screenshot shows a GitHub pull request titled "ncarkaci Merge pull request #7 from keyiflerolsun/master". It lists several files and their commit messages:

File	Commit Message	Time
sözlükler	Birden fazla sözlüğü tek bir sözlükte birleştirme kodu tdk_sözlük dos...	2 years ago
.gitignore	gitignore file added.	2 years ago
Birleştirilmiş_Sözlük_Kelime_Listesi.txt	TDK ve Zemberek birleştirilmiş dosya eklendi.	2 years ago
README.md	Bazı son düzenlemeler yapıldı.	5 years ago
TDK_Sözlük_Kelime_Listesi.txt	Kelime dosyası eklendi.	5 years ago
istatistik.py	Küçük Dokunuşlar..	11 months ago
kelime_getir.py	Küçük Dokunuşlar..	11 months ago
n-gram.py	Küçük Dokunuşlar..	11 months ago
ortak_kelimeler.txt	Ortak ve farkları bulan kod tdk sözlük içine taşındı.	2 years ago
tdk_fark_zemberek.txt	Ortak ve farkları bulan kod tdk sözlük içine taşındı.	2 years ago
tdk_sozluk.py	Küçük Dokunuşlar..	11 months ago
zemberek_fark_tdk.txt	Ortak ve farkları bulan kod tdk sözlük içine taşındı.	2 years ago

Şekil 1

Şekil 1'de görüldüğü gibi Necmettin Çarkacı isimli yazılım geliştiricisinin daha evvel TDK'dan ve zemberek kütüphanesinden topladığı sözlükler ve bu sözlüklerin ortak kelimeleri mevcuttur. Ortak kelimelerin ayrıca bulunmasıyla kuvvetli bir sözlükten bahsedilebilir. Dolayısıyla buradan faydalanılabilir ve kaynak olarak gösterilebilir.

Code 1

```
>>> ortak_kelimeler = open('ortak_kelimeler.txt', mode = 'r')
```

Code 1'de görüldüğü şekliyle dosya okunduğu taktirde, Türkçe karakterlerin ekrana bastırılması ve kullanılması durumlarında bilinmeyen karakter ifadelerine denk gelinebilir. Bu sebeple özellikle dosya açılırken farklı bir kod çözücü ile açılması gerekmektedir.

Code 2

```
>>> ortak_kelimeler = open('ortak_kelimeler.txt', encoding = 'utf-8', mode = 'r')
>>> ortak_kelimeler = ortak_kelimeler.read()
>>> ortak_kelimeler = ortak_kelimeler.split("\n")
>>> len(ortak_kelimeler)
41337
```

Code 2’de dosya doğru bir kod çözümleme ile açılıp kelimeler bölünerek liste haline getirilmiştir. Toplamda, 41337 kelime mevcuttur.

Code 3

```
>>> sozluk = open("sozlukler/z.txt", encoding = 'utf-8', mode = 'w')
>>> for i in range(40856,41337,1):
>>>     sozluk.write(ortak_kelimeler[i] + "\n")
>>> sozluk.close()
```

Code 3’te genel ve ortak bir sözlüğün partition işlemi yapılmıştır. A’dan Z’ye tüm alfabeyle ait sözcükler birbirinden ayrılarak, harfe ait bir text dosyası içerisine kaydedilmiştir.

Code 4

```
>>> b_harfi = open("sozluk/b.txt", encoding = 'utf-8', mode = 'r')
>>> b_harfi_ile_basyalan_kelimeler = b_harfi.read()
>>> b_harfi_ile_basyalan_kelimeler.split()
['baba',
 'babaanne',
 'babaannelik',
 'babacanca',
 ...,
 'büzüşmek',
 'büzüşük']
>>> len(b_harfi_ile_basyalan_kelimeler)
27705
```

Code 4’te, rastgele seçilen “b” harfinin text dosyasının çıktısı görülmektedir. Bu şekilde bir partition yapılması, dökümandaki terimlerin temizlenmesinden sonra kontrol yapısını kolaylaştıracaktır. Ayrıca performanslı bir çıktı elde etmeye sebebiyet vereceği kesindir.

Parçalı bir sözlüğün elde edilmesi her bir kelime başlangıcındaki harfin elde edilmesi gerektiği sonucunu doğurmaktadır. Dolayısıyla bu durumda ilk harfin elde edilmesi gibi ek durumların performansı ister istemez düşürecektir.

REFERANS VE KAYNAKÇA

- [1] Dil bilgisi için klavuz TDK. Link için [tıklayınız](#). (Güvenlidir.)
- [2] Yemeksepeti gizlilik politikası. Link için [tıklayınız](#). (Güvenlidir.)
- [3] Yemeksepeti kişisel verilerin korunumu. Link için [tıklayınız](#). (Güvenlidir.)
- [4] BeautifulSoup kütüphanesi dökümanları. Link için [tıklayınız](#).

EKLER

Bitirme Projesi 2'ye ait doküman, program kodu, haftalık rapor ve ek bilgilerin paylaşıldığı github linki için [tıklayınız](#). (Güvenlidir.)