

LAB-8

Burak uzun 210911131

Not:database işine kafam yatmadığı için database ye kaydedemedim onun dışında bir sorun yok.

1-soru 1:

Özet:

öncelikl get soup fonksiyonundan başlamak istiyorum.bu fonksiyon bize verilen linkteki sitenin html kodlarını doğru bir biçimde karşımıza çıkarmayı sağlar.

Çözüm mimarisi:

Sorunu çözümü için fonksiyona link gönderiyoruz.burda rutin derleme işleminden farklı bir şey yapmıyorum o yüzden kısa kesiyorum.

Çözüm fonksiyonu:

```
def get_soup(self,webpage):  
    """girdi olarak verilen sayfanın linkindeki html kodlarını  
    güzelleştirip return eder  
  
    Args:  
  
    input_file_name: webpage  
    output_file_name: soup  
  
    Returns:  
  
    soup  
  
    """  
    self.gelensayfalinki=webpage  
    self.c = urllib2.urlopen(self.gelensayfalinki)  
    self.metin = self.c.read()  
    self.soup = BeautifulSoup(self.metin,features="html.parser")  
    self.soup.prettify()  
  
    return self.soup
```

2- soru2:

Özet:

Get_Categories fonksiyonu , bize verilen main fonksiyonun side bar da olan kategorilerini ayrıştırma imkanı sunuyor. Bunları bir sözlüğe hapsedebiliyor.

Çözüm mimarisi:

Öncelikle main linki yani sitenin ana sayfasının linkini mainsite değişkenine tanımladım. Ardından bu site linkini get_soup fonksiyonuna gönderdim. Bize return edilen soup üzerinden navlistlerin araştırmasını yaptım ve bize a_listesini verdi. Bu listenin içinde,htmllerinden arıştırılmamış link ve kategoriler var. 42. Satırda bir for döngüsü açtım ve bu listenin içinde htmlleri arındırmak için döndüm. Bu for döngüsü kısaca,htmllerde arındırıp,çok temiz bir ikili kategori,link listesi veriyor.

Çözüm fonksiyonu:

```
def get_categories(self):
    """main sitedeki kategorileri ve kategorilerin linklerini sözlük
    olarak dışa verir

    Args:

    input_file_name: none

    Returns:

    kategorisözlüğü

    """
    self.mainsite="http://books.toscrape.com/index.html"

    self.soup=self.get_soup(self.mainsite)

    self.navlistler = self.soup.find_all(class_='nav nav-list')
    self.a_listesi = self.navlistler[0].findChildren("a" , recursive=True)

    self.kategorisozlugu={}
    for i in range(1,len(self.a_listesi)):

        self.stripcategories=self.a_listesi[i].string.strip()
        self.link=self.a_listesi[i].get("href")
        self.link_birlesmis = urljoin(self.mainsite, self.link)

        self.kategorisozlugu[self.stripcategories]=self.link_birlesmis
    return self.kategorisozlugu
```

3-soru 3:

Özet:

Verilen kategorilerin içindeki kitap bilgilerini çeker ve sözlüğe atar.

Çözüm mimarisi:

Fonksiyon argüman alan bir fonksiyon önce soup ve link adında değişkenler alır. Bunlardan soup olan bize,kitap bilgilerini çekebilmemiz için temiz bir çıktı sağlıyor. Link ise kitapların linklerini ilerde join etmek için gerekli.

Öncelikle sitedeki h3 etiketlerini bulum. Bu etiketin içiinde kitap isimleri var. Değer değişkeninde,kitabın fiyatı var.nesneler adında değişken oluşturdum bu nesneyi oluşturmamın sebebi for döngüsünün içinde oluşturduğum sözlüğü doğrudan listeye append etmek. For döngüsünün içine geçelim kitaplar sözlüğü oluşturdum bu oluşumun peşinden değişken isimlerinden anlaşıldığı gibi kitao ismi fiyatı ve href yani link i geliyor. Ardınan bu linki joinliyoruz yani birleştiriyoruz ve kitabın linkini ortaya çıkartıyoruz. Ardınan bunlardan sözlükler sözlüğü oluşturup bir listeye append ediyoruz.

Çözüm fonksiyonu:

```
def get_prices_stars(self,soup,link):
    """verilen kategorilerin içindeki kitapların bilgilerini çeker ve
    dışarıya atar

    Args:

    input_file_name: soup,link

    Returns:

    nesneler

    """
    self.link=link
    self.h3=soup.find_all("h3")
    self.deger=soup.find_all(class_="price_color")

    self.nesneler=[]
    for i in range(len(self.h3)):
        self.kitaplarSoz={}
        self.kitapAdi=self.h3[i].string.strip().strip()
        self.KitapFiyati=self.deger[i].string
        self.a_listesi = self.h3[i].findChildren("a" , recursive=True)
        self.a_linki=self.a_listesi[0].get("href")
        self.joinlink=urljoin("{}".format(self.link),self.a_linki)

        self.kitaplarSoz["Name"]=self.kitapAdi
        self.kitaplarSoz["Price"]=self.KitapFiyati
        self.kitaplarSoz["Url"]=self.joinlink

        self.nesneler.append(self.kitaplarSoz)

    return self.nesneler
```

4-soru 4:

Özet:

Get_Categories içindeki kategoriler ve linkler içinde for döngüsüyle gezer ve bunları get_prices_Star içine gönderir.

Çözüm mimarisi:

Bu işlemin takibini yapabilmek için bir tane progress bar ekledim , ilk 3 değişken progress bar ile ilgili.

107. satırda genel bir sözlük oluşturunca bu sözlük bütün kategorilerin bilgilerini barındıracak. For döngüsünün içinde kategorileri geziyor link ve kategorileri temizlemek için get_soup a gönderiyor. Ardından souplusayfa değişkenini prices_stars a gönderiyoruz fonksiyondan işlenen veriler genel sözlüğe kat=kategori kat1:kitap bilgileri olarak kaydediliyor. Bu satırdan sonraki satır progress bar içindir.

Çözüm fonksiyonu:

```
def parse(self):
    """kategorileri tarar ve get_prices_stars fonksiyonuna gönderir ve
    bütün
    kategorideki kitapların bilgilerini bir sözlükte depolar

    Args:

    nline

    Returns:

    genelsozluk

    """

    total = 50 # total number to reach
    bar_length = 34 # should be less than 100
    i=0

    self.genelsozluk={}
    self.x=self.get_categories()
    for kat,lik in self.x.items():
        self.souplusayfa=self.get_soup(lik)
        self.KAT1=self.get_prices_stars(self.souplusayfa,lik)
        self.genelsozluk[kat]=self.KAT1
        ###progress bar
        i=i+1
        percent = 100.0*i/total
        sys.stdout.write('\r')
        sys.stdout.write("Completed: [{:}] {:>3}%"
```

```

        .format('='*int(percent/(100.0/bar_length)),
                bar_length, int(percent)))

    sys.stdout.flush()
    time.sleep(0.002)
    ###progress bar
    return self.genelsozluk

```

Test fonksiyonu:

```

from BURAK_UZUN_211911131_LAB8 import WebScraper

def main():
    a=WebScraper()

    #yorumu kaldırırsak sadece kategori ve link sözlüğü verir.
    #print(a.get_categories())

    #yorummu kaldırırsak bütün kitap bilgileri ve kategorileri verir.
    #print(a.parse())

main()

```

test çıktısı:

```

{'Travel': 'http://books.toscrape.com/catalogue/category/books/travel_2/index.html', 'Mystery': 'http://books.toscrape.com/catalogue/category/books/mystery_3/index.html', 'Historical Fiction': 'http://books.toscrape.com/catalogue/category/books/historical-fiction_4/index.html', 'Sequential Art': 'http://books.toscrape.com/catalogue/category/books/sequential-art_5/index.html', 'Classics': 'http://books.toscrape.com/catalogue/category/books/classics_6/index.html', 'Philosophy': 'http://books.toscrape.com/catalogue/category/books/philosophy_7/index.html', 'Romance': 'http://books.toscrape.com/catalogue/category/book/romance_8/index.html', 'Womens Fiction': 'http://books.toscrape.com/catalogue/category/books/womens-fiction_9/index.html', 'Fiction': 'http://books.toscrape.com/catalogue/category/books/fiction_10/index.html', 'Childrens': 'http://books.toscrape.com/catalogue/category/books/childrens_11/index.html', 'Religion': 'http://books.toscrape.com/catalogue/category/books/religion_12/index.html', 'Nonfiction': 'http://books.toscrape.com/catalogue/category/books/nonfiction_13/index.html', 'Music': 'http://books.toscrape.com/catalogue/category/books/music_14/index.html', 'Default': 'http://books.toscrape.com/catalogue/category/books/default_15/index.html', 'Science Fiction': 'http://books.toscrape.com/catalogue/category/books/science-fiction_16/index.html', 'Sports and Games': 'http://books.toscrape.com/catalogue/category/books/sports-and-games_17/index.html', 'Add a comment': 'http://books.toscrape.com/catalogue/category/books/add-a-comment_18/index.html', 'Fantasy': 'http://books.toscrape.com/catalogue/category/books/fantasy_19/index.html', 'New Adult': 'http://books.toscrape.com/catalogue/category/books/new-adult_20/index.html', 'Young Adult': 'http://books.toscrape.com/catalogue/category/books/young-adult_21/index.html', 'Science': 'http://books.toscrape.com/catalogue/category/books/science_22/index.html', 'Poetry': 'http://books.toscrape.com/catalogue/category/books/poetry_23/index.html', 'Paranormal': 'http://books.toscrape.com/catalogue/category/books/paranormal_24/index.html', 'Art': 'http://books.toscrape.com/catalogue/category/books/art_25/index.html', 'Psychology': 'http://books.toscrape.com/catalogue/category/books/psychology_26/index.html', 'Autobiography': 'http://books.toscrape.com/catalogue/category/books/autobiography_27/index.html', 'Parenting': 'http://books.toscrape.com/catalogue/category/books/parenting_28/index.html', 'Adult Fiction': 'http://books.toscrape.com/catalogue/category/books/adult-fiction_29/index.html', 'Humor': 'http://books.toscrape.com/catalogue/category/books/humor_30/index.html', 'Horror': 'http://books.toscrape.com/catalogue/category/books/horror_31/index.html', 'History': 'http://books.toscrape.com/catalogue/category/books/history_32/index.html', 'Food and Drink': 'http://books.toscrape.com/catalogue/category/books/food-and-drink_33/index.html', 'Christian Fiction': 'http://books.toscrape.com/catalogue/category/books/christian-fiction_34/index.html', 'Business': 'http://books.toscrape.com/catalogue/category/books/business_35/index.html', 'Biography': 'http://books.toscrape.com/catalogue/category/books/biography_36/index.html', 'Thriller': 'http://books.toscrape.com/catalogue/category/books/thriller_37/index.html', 'Contemporary': 'http://books.toscrape.com/catalogue/category/books/contemporary_38/index.html', 'Spirituality': 'http://books.toscrape.com/catalogue/category/books/spirituality_39/index.html', 'Academic': 'http://books.toscrape.com/catalogue/category/books/academic_40/index.html', 'Self Help': 'http://books.toscrape.com/catalogue/category/books/self-help_41/index.html', 'Historical': 'http://books.toscrape.com/catalogue/category/books/historical_42/index.html', 'Christian': 'http://books.toscrape.com/catalogue/category/books/christian_43/index.html', 'Suspense': 'http://books.toscrape.com/catalogue/category/books/suspense_44/index.html', 'Short Stories': 'http://books.toscrape.com/catalogue/category/books/short-stories_45/index.html', 'Novels': 'http://books.toscrape.com/catalogue/category/books/novels_46/index.html', 'Health': 'http://books.toscrape.com/catalogue/category/books/health_47/index.html', 'Politics': 'http://books.toscrape.com/catalogue/category/books/politics_48/index.html', 'Cultural': 'http://books.toscrape.com/catalogue/category/books/cultural_49/index.html', 'Erotica': 'http://books.toscrape.com/catalogue/category/books/erotica_50/index.html', 'Crime': 'http://books.toscrape.com/catalogue/category/books/crime_51/index.html'}

```

Soru1 test çıktısı

```

Completed: [=====] 22%

```

Progress bar

```
Completed: [=====] 100% Travel: [{ 'Name': 'It's Only the Himalayas', 'Price': '£45.17', 'Url': 'http://books.toscrape.com/catalogue/its-only-the-himalayas_981/index.html'}, {'Name': 'Full Moon over Noah's ...', 'Price': '£49.43', 'Url': 'http://books.toscrape.com/catalogue/full-moon-over-noahs-ark-an-odyssey-to-mount-ararat-and-beyond_811/index.html'}, {'Name': 'See America: A Celebration ...', 'Price': '£48.87', 'Url': 'http://books.toscrape.com/catalogue/see-america-a-celebration-of-our-national-parks-treasured-sites_732/index.html'}, {'Name': 'Vagabonding: An Uncommon Guide ...', 'Price': '£36.94', 'Url': 'http://books.toscrape.com/catalogue/vagabonding-an-uncommon-guide-to-the-art-of-long-term-world-travel_552/index.html'}, {'Name': 'Under the Tuscan Sun', 'Price': '£37.33', 'Url': 'http://books.toscrape.com/catalogue/under-the-tuscan-sun_504/index.html'}, {'Name': 'A Summer In Europe', 'Price': '£44.34', 'Url': 'http://books.toscrape.com/catalogue/a-summer-in-europe_458/index.html'}, {'Name': 'The Great Railway Bazaar', 'Price': '£30.54', 'Url': 'http://books.toscrape.com/catalogue/the-great-railway-bazaar_446/index.html'}, {'Name': 'A Year in Provence ...', 'Price': '£56.88', 'Url': 'http://books.toscrape.com/catalogue/a-year-in-provence-provence-1_421/index.html'}, {'Name': 'The Road to Little ...', 'Price': '£23.21', 'Url': 'http://books.toscrape.com/catalogue/the-road-to-little-dribbling-adventures-of-an-american-in-britain-notes-from-a-small-island-2_277/index.html'}, {'Name': 'Neither Here nor There: ...', 'Price': '£38.95', 'Url': 'http://books.toscrape.com/catalogue/neither-here-nor-there-travels-in-europe_198/index.html'}, {'Name': '1,000 Places to See ...', 'Price': '£26.08', 'Url': 'http://books.toscrape.com/catalogue/1000-places-to-see-before-you-die_1/index.html'}], 'Mystery': [{ 'Name': 'Sharp Objects', 'Price': '£47.82', 'Url': 'http://books.toscrape.com/catalogue/sharp-objects_997/index.html'}, {'Name': 'In a Dark, Dark ...', 'Price': '£19.63', 'Url': 'http://books.toscrape.com/catalogue/in-a-dark-dark-wood_963/index.html'}, {'Name': 'The Past Never Ends', 'Price': '£56.50', 'Url': 'http://books.toscrape.com/catalogue/the-past-never-ends_942/index.html'}, {'Name': 'A Murder in Time', 'Price': '£16.64', 'Url': 'http://books.toscrape.com/catalogue/a-murder-in-time_877/index.html'}, {'Name': 'The Murder of Roger ...', 'Price': '£44.10', 'Url': 'http://books.toscrape.com/catalogue/the-murder-of-roger-ackroyd-hercule-poirot-4_852/index.html'}, {'Name': 'The Last Mile (Amos ...', 'Price': '£54.21', 'Url': 'http://books.toscrape.com/catalogue/the-last-mile-amos-decker-2_754/index.html'}, {'Name': 'That Darkness (Gardiner and ...', 'Price': '£13.92', 'Url': 'http://books.toscrape.com/catalogue/that-darkness-gardiner-and-renner-1_743/index.html'}, {'Name': 'Tastes Like Fear (DI ...', 'Price': '£10.69', 'Url': 'http://books.toscrape.com/catalogue/tastes-like-fear-di-marnie-rome-3_742/index.html'}, {'Name': 'A Time of Torment ...', 'Price': '£48.35', 'Url': 'http://books.toscrape.com/catalogue/a-time-of-torment-charlie-parker-14_657/index.html'}, {'Name': 'A Study in Scarlet ...', 'Price': '£16.73', 'Url': 'http://books.toscrape.com/catalogue/a-study-in-sca
```

2-3-4. soruların ortak çıktısı

Ek dosyalar:

BURAK_UZUN_211911131_LAB8.py

BURAK_UZUN_211911131_LAB8.pdf