# Yaz. Lab – 1 Proje Ara Raporu

Buse Çakal (221307084)

Betül Erdik (211307033)

**1. Proje Amacı**

Bu proje, belirli bir web sayfasında bulunan .mp3 formatındaki dosyaların tespit edilmesi, taranması ve kullanıcı sistemine indirilmesini sağlamaktadır. Projede Selenium ve BeautifulSoup gibi Python kütüphaneleri kullanılmıştır.

**2.Kullanılan Kütüphaneler**

**1.Selenium:** Web sitelerinde kullanıcı hareketlerini taklit ederek tarayıcı üzerinden işlem yapılmasını sağlar. Bu projede, web sitesini açmak ve HTML içeriğini almak için kullanılmıştır.

**2.** **BeautifulSoup:** **Web sayfasının HTML içeriğini parse ederek yapılandırılmış bir şekilde işlem yapmaya olanak tanır.**

**3.** **Requests: HTTP istekleri yaparak web sayfalarından veya dosyalardan veri çekmek için kullanılır. Projede, MP3 dosyalarını indirmek için kullanılmıştır.**

**4.** **Webdriver Manager: Selenium için gereken ChromeDriver veya diğer tarayıcı sürücülerini dinamik olarak yükler ve kurar.**

**5.** **os:** İşletim sistemi ile etkileşim kurarak klasör oluşturma ve dosya yollarını yönetmek için kullanılmıştır.

**6.** **urllib.parse:** Relatif ve mutlak URL'leri işlemek için kullanılır.

**service\_account.json:** Google Drive API ile iletişim için gerekli yetkilendirme bilgilerini içerir. Bu dosya, Google Cloud Platform'dan indirilmiştir.

**main.py:** Projenin ana işlevselliğini içeren Python kodudur. Dosyaların indirilmesi ve yüklenmesi işlemlerini gerçekleştirir.

**3. Google Cloud API Ayarları**

İlk olarak “Google Cloud Console” içerisinde yeni bir proje oluşturuldu. Proje oluştuğunda bizi karşılayan sayfa içerisindeki "API ve Hizmetler" sekmesine gidildi. "Kitaplık" seçeneğinden Google Drive API etkinleştirildi.

"Kimlik Bilgileri" sekmesinden "Hizmet Hesabı" oluşturuldu. Hizmet hesabına uygun bir rol (ör. Drive File Access) atandı. İstenirse farklı roller de atamak mümkün veya oluşturulan hizmet hesabını farklı rollerle başka kişilerle paylaşıma açmak da mümkün.

Böylece hizmet hesabı için bir json dosyası oluşturuldu ve bilgisayara indirildi. Bu dosya,” service\_account.json” adıyla proje dizinine kaydedildi.

Sonrasında proje içinde kullanılacak Google Drive klasörü oluşturuldu. Hizmet hesabı e-posta adresi (JSON dosyasında yer alır) bu klasöre düzenleme yetkisiyle davet edildi. Bu şekilde drive dosyasına hizmet hesabına erişim izni verilmiş oldu.

4. Kod Açıklamaları

from selenium import webdriver

from selenium.webdriver.common.by import By

from selenium.webdriver.chrome.service import Service

from webdriver\_manager.chrome import ChromeDriverManager

import requests

from bs4 import BeautifulSoup

import os

from urllib.parse import urljoin

*Bu segmentte ilk olarak proje için gerekli tüm kütüphaneler import edildi.*

download\_folder = "mp3\_files"

os.makedirs(download\_folder, exist\_ok=True)

*MP3 dosyalarını depolamak için mp3\_files adlı bir klasör oluşturulur. Eğer klasör zaten mevcutsa hata almamak için exist\_ok=True parametresi kullanılmıştır.*

options = webdriver.ChromeOptions()

options.headless = True # Headless modda çalıştır

driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), options=options)

*Selenium WebDriver, tarayıcıyı otomatik olarak çalıştırır. headless=True ile tarayıcı arka planda çalıştırılır.*

driver.get(base\_url)

html\_content = driver.page\_source

*Belirtilen URL'ye gidilir ve sayfanın HTML içeriği alınır.*

soup = BeautifulSoup(html\_content, 'html.parser')

mp3\_links = [urljoin(base\_url, link['href']) for link in soup.find\_all('a', href=True) if link['href'].endswith('.mp3')]

*Sayfa içeriği BeautifulSoup kullanılarak işlenir. .mp3 ile biten tüm bağlantılar listeye eklenir. Relatif URL'ler urljoin ile mutlak URL'ye dönüştürülür.*

for mp3\_link in mp3\_links:

try:

response = requests.get(mp3\_link, stream=True)

response.raise\_for\_status()

file\_name = os.path.join(download\_folder, os.path.basename(mp3\_link))

with open(file\_name, 'wb') as file:

for chunk in response.iter\_content(chunk\_size=8192):

file.write(chunk)

print(f"Başarıyla indirildi: {file\_name}")

except Exception as e:

print(f"İndirme sırasında hata oluştu: {e}")

*Her MP3 bağlantısı için requests.get ile HTTP isteği yapılır. Gelen içerik belirli bir tampon boyutunda yazılarak dosyaya kaydedilir. Bu yöntem, büyük dosyaları bellekte tutmadan indirmeyi sağlar.*

driver.quit()

*İşlemler tamamlandıktan sonra Selenium tarafından açılan tarayıcı kapatılır.*

Proje çalıştırıldığında:

Belirtilen URL'den .mp3 dosyaları tespit edilir.

Dosyalar mp3\_files klasörüne indirilir.

İndirme işlemi sırasında herhangi bir hata meydana gelirse konsola ilgili hata mesajı yazdırılır.