

DEÜ Fen Fakültesi
Bilgisayar Bilimleri Bölümü

BİL3013 Veri Madenciliğine Giriş

Ödev 1. Web kazıma tekniği ile veri toplama

Kodların ve Veri setlerinin Kaggle linki:

<https://www.kaggle.com/code/efendinasiboglu/notebook...>

Grup 1

Öğrenci: Kemal SUNAL

Öğretim Üyesi: Prof. Dr. Efendi NASİBOĞLU

İzmir 2024

I. ÖDEVİN TANIMI

GRUP 1) sahibinden.com sitesinden 2. el **Fiat** markalı arabaların **İlan tarihi, İlan metni, İli, İlçesi, Modeli, Rengi, Yılı, Vites türü, Yakıt türü, Fiyatı** gibi özelliklerini web kazıma yöntemleriyle çekerek bir **Excel** dosyasında depolayan **Python** uygulamasını geliştirin. Veriler çekilen sayfalarda **sayfalama** veya **scrolling** işlemleri **otomatik** yapılacaktır. **En az 500** örnek veri toplanacaktır.

Yapılması gerekenler:

1. Her öğrenci **kaggle.com** sitesinde kişisel hesap açmalı.
2. Geliştirdikleri **program kodlarını ve oluşturdukları veri setini** bu sitede public olarak diğer kullanıcılarla veya isterlerse dersin öğretim üyesi olarak sadece benimle **paylaşmalı**.
3. Herkes **kışisel** Ödev Raporunu **Word** dosyası olarak **Sakai** sisteminde **Ödev** sekmesine yüklemeli. Raporun kapak sayfasında erişim için öğrencinin **Kaggle linkini** de yazmalı.

II. KULLANILAN YÖNTEMLER VE TEKNOLOJİLER

BeautifulSoup. Bu bir Python kütüphanesidir. Web sitelerinin html şeklinde ele alınarak, içerisinden gereken bilgilerin çekilmesine olanak sağlıyor. Kütüphanenin indirilmesi ve programda kullanılması için aşağıdaki komutlar kullanılabilir:

```
!pip install beautifulsoup4
import requests
from bs4 import BeautifulSoup
```

Belli siteden veri çekmek için aşağıdaki komutlar kullanılabilir:

```
import requests
from bs4 import BeautifulSoup
urunLink = "https://www.hepsiburada.com/iphone-se-64-gb-p-HBV00000SXR45"
site = requests.get(urunLink)
siteIcerigi = BeautifulSoup(site.content, "lxml")
```

....

...

Selenium. Bu bir Python kütüphanesidir. Web sitelerinin üzerinde gezinme, tıklama vb. işlemleri otomatikleştirmek için kullanılabilir....

...

...

III. UYGULAMA

3.1. Veri toplama için sahibinde.com sitesi kullanılmıştır. Verileri çekerken, önce gereken sitenin bilgileri aşağıdaki komutlar aracılığıyla çekildi:

....

....

...

3.2. Uygulama çalışma sonucu oluşan ekran görüntüleri aşağıdaki Şekil 1-3’de verilmiştir.

...

...

IV. DEĞERLENDİRME

Burada ne gibi zorluklarla karşılaştınız ve onları nasıl çözdünüz gibi bilgiler verilebilir.

...

Kaynaklar

1. Ali Eren Ekinci. Python’da veri çekmek – Requests ve BeautifulSoup4 rehberi. Erişim tarihi: 9.10.2024. <https://www.alierenekinci.com/uncategorized/pythonda-veri-cekme-request-ve-beautifulsoup4-rehberi/>
2. Yusuf Talha ARABACI. Hackerlar için Python: selenium. Erişim tarihi: 9.10.2024. <https://medium.com/@yusufarbc/hackerlar-i%C3%A7in-python-selenium-d398454c3b9b>
3. Makale...
4. Kitap...
5. vs