

KOCAELİ ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
YAZILIM LAB. I- I. Proje

PROJE TESLİM TARİHİ: 23.10.2022

Web Scraping E-ticaret

Trendyol, Hepsiburada, n11, Amazon, gibi e ticaret sitelerinden istenilen ürüne ait bilgilerin yer aldığı bir veritabanı oluşturulacak ve bu veritabanından istenilen bilgileri web sayfası üzerinde gösterilecektir.

Amaç:

1. Web scraping ile farklı alışveriş sitelerinden istenilen ürün hakkında bilgi elde etmek.
2. Bir uygulama içerisinde istenilen özellikteki ürünlerin filtrelenmesi ve sıralanması becerisini geliştirmek,
3. Dinamik özelliklere sahip bir program geliştirmek
4. Web kodlama hakkında bilgi ve beceriye sahip olmak

Programlama Dili: C, C#, JAVA, Python. vb.

Web Scraping:

1. 4 adet E-Ticaret (Trendyol, Hepsiburada, n11, Amazon) sitesi üzerinden web scraping kullanılarak bilgisayar(notebook) kategorisine ait bilgiler alınacaktır. Ayrıca E-ticaret sitelerine benzer yapıda kendiniz bir web sitesi oluşturacak (sadece bilgisayar(notebook) kategorisi olacak) ve web scraping kullanarak sitedeki bilgiler alınacaktır.
2. Web Scraping kullanılırken siteye ait html bilgiler veya siteye request isteği ile bilgiler alınmalıdır.

Veritabanı:

1. Web scraping ile elde edilen veriler veritabanına kaydedilecektir.
2. Kullanılacak veritabanı proje uygun olarak herhangi bir tanesini kullanabilirsiniz (MongoDB gibi veri tabanları kullanılması artı puan verilecektir.)
3. Veritabanında aşağıdaki tabloda belirtilen özellik bilgilerinin kullanılması gerekmektedir. Veritabanındaki tabloların normalizasyon kurallarına uygun olacak şekilde tasarlanması ve organize edilmelidir.

Marka	Model Adı	Model No	İşletim sistemi	İşlemci tipi	İşlemci nesli	Ram	Disk boyutu	Disk türü	Ekran boyutu	Puanı	Fiyat	Site ismi
HP	250G8	34N99ES03	Freedos	i7	11	16	256	SSD	15.6	4.0	15.699	Hepsiburada
...
...

Web Sayfası:

1. Farklı sitelerden çekilen ürün bilgilerinin gösterilmesi için bir web sayfası oluşturmanız beklenmektedir. Örnek web sayfası aşağıdaki resimde görülmektedir.
2. Bu sayfa ilk açıldığında veritabanında yer alan tüm kayıtlar listelenmelidir.
3. Listeleme işleminde ürün bilgilerinin yer aldığı ana başlık bulunmalıdır. Herhangi bir ürün başlığına tıklandığında ürün hakkındaki bilgilerin (fiyat ve ürün özellikleri) olduğu sayfaya yönlendirilecektir..
4. Listeleme işleminde ürüne ait en düşük 3 fiyat ve bunların ait olduğu E- ticaret sitelerinin linkleri isim ve logolarıyla birlikte yer alacaktır. İsim veya logonun üzerine tıklayınca ürünün bulunduğu E-ticaret sitesinin sayfası açılacaktır.

Markası

- ☐ Dell
- ☐ Huawei
- ☐ Apple
- ☐ Asus
- ☐ Monster
- ☐ MSI
- ☐ Acer
- ☐ HP
- ☐ Lenovo
- ☐ Honor
- ☐ Casper
- ☐ iXtech
- ☐ Thomson
- ☐ Hometech

Çok Satanlar	En Düşük Fiyat	En Yüksek Fiyat	En Yüksek Puan
<div> </div> <div> <p>Dell G15 5511 G55112306U i5-11260H 16 GB 512 GB SSD RTX 3050 15.6\"</p> <p>EN Ücuz 17.500,00 TL Ücretsiz kargo </p> <p>17.999,00 TL Ücretsiz kargo </p> <p>18.498,99 TL +18,99 TL kargo </p> </div>			

5. Web sayfasından herhangi bir ürünle ilgili dinamik arama işlemi yapılabilmelidir. **Bu arama alanında bir ürünün isim, model ya da bulunduğu E-ticaret sitesine göre yapılmalıdır. Ayrıca Arama sırasında yazım yanlışı olması durumunda sistem düzeltilmiş öneride bulunmalıdır.** Örnek: hepsi burada -- yazımı düzeltilmiş sonuçları görüyorsunuz: Hepsiburada şeklinde olmalıdır.
6. Web sayfasında **dinamik filtreleme işlemi** de ayrıca yer almalıdır. Filtreleme işlemi veritabanında yer alan ürünün tüm özellikleri için uygulanabilir olmalıdır. Yukarıdaki görselde e-ticaret sitelerinde bulunan laptop ürünleri markasına göre örnek olarak filtrelenmiştir.
7. Web sayfasında en düşük fiyattan en yüksek fiyata, en yüksek fiyattan en düşük fiyata ve en yüksek puandan en düşük fiyata şeklinde sıralama işlemleri yapılabilmelidir. Laptop bilgilerinin yer aldığı alanda fiyat bilgisi için de aynı durum geçerli olacaktır.

Duplicate :

1. Scraping ile çekilen bilgiler güncellenmesi gerekmektedir. Güncel bilgiler veritabanında kaydedilmelidir.

2. Kaydedilen veriler ile önceki kayıtlı veriler benzer bilgiler içermektedir. Bundan dolayı Duplicate oluşmaması için veritabanında kontrol işlemi yapılmalı ve veritabanı ile web sayfasında sadece güncel bilgiler yer almalıdır.
3. Veritabanında Near Duplicate kontrolü yapılacaktır. Veritabanında Benzer isimlerde kaydedilmiş birden fazla kayıt olabilir. Bu durum aynı isimlerde farklı kayıtlar olduğu anlamına gelmektedir. Örneğin bir web sitesinden ürünün işletim sistemi bilgisi Dos olarak diğerinden ise FreeDos olarak kayıt edildiğini düşünelim. Bu iki durum aynı bilgiyi temsil etmektedir. Bundan dolayı Dos olanı diğer unique bir etiket altında temsil edilmelidir.

Örnek Web Site Linkleri

- <https://www.cimri.com/dizustu-bilgisayar>
- <https://www.akakce.com/laptop-notebook.html>

İsterler:

1. Kendinize ait bir E-ticaret sitesi oluşturmalsınız ve bu site Admin tarafından güncellenebilir olmalıdır. Sitede yapılacak değişikliklerde (
 - Bir notebook ürünün fiyatının değiştirilmesi
 - Notebook ürünü ait bir kaydın kaldırılması
 - Yeni notebook ürünün eklenmesi
 - Notebook ürünün puanında değişiklik yapılması
 - Ürün bilgilerinin güncellenmesi) bilgilerin web scraping ile çekilmesi gerekmektedir. Alınan tüm bu değişiklikler farklı sitelerdeki ürünlerin listelendiği diğer web sitesinde anlık olarak gösterilmelidir.
2. Ürün bilgisinin gösterildiği web sayfasında tüm notebook ürünlerinin listelenmesi ve ürün bilgilerinin yer aldığı ana başlıklarının oluşturulması gerekmektedir. Bu başlığa tıklayınca ürün hakkında bilgilerin (fiyat bilgileri dahil) olduğu sayfa açılmalıdır. Ayrıca ürünlerin listelendiği sayfadaki E-ticaret linki üzerine tıklayınca ürünün bu E-ticaret sitesindeki sayfasına kullanıcı yönlendirilmelidir.
3. a) Dinamik arama ile (yukarıda istenilen tüm arama kriterleri dahil olacak şekilde) Asus ..vb modele göre arama işleminin yapılması

b) Dinamik arama ile (yukarıda istenilen tüm arama kriterleri dahil olacak şekilde) hepsiburada ..vb E-ticaret sitelerinin aranması
4. a) Ürün bilgilerinin gösterildiği web sayfası üzerinde Filtreleme işleminin yapılması

b) Fiyat bilgisine göre küçükten büyüğe veya büyükten küçüğe sıralama ayrıca puanı yüksek olan ürüne göre sıralama işleminin yapılması
5. Sizin oluşturacağınız E- Ticaret sitesine kayıtlı olan bir ürünün aynı bilgileri ile kayıt edilmeli ve veritabanında Duplicate Kontrolü yapılmalıdır.
6. Sizin oluşturacağınız E- Ticaret sitesinde iki farklı kayıt işleminin yapılması ve Near Duplicate kontrolünün yapılması

Proje Teslimi

- Rapor IEEE formatında (önceki yıllarda verilen formatta) 4 sayfa, akış diyagramı veya yalancı kod içeren, özet, giriş, yöntem, deneysel sonuçlar, sonuç ve kaynakça bölümünden oluşmalıdır. Raporda kullanılan algoritma açıklanmalı ve algoritmanın kaba kodu yazılmalıdır.
- Dersin takibi projenin teslimi dahil edestek.kocaeli.edu.tr sistemi üzerinden yapılacaktır. edestek.kocaeli.edu.tr sitesinde belirtilen tarihten sonra getirilen projeler kabul edilmeyecektir.
- Proje ile ilgili sorular edestek.kocaeli.edu.tr sitesindeki forum üzerinden Arş. Gör. Emin ÖLMEZ veya Arş. Gör. Muhammed Ahmet DEMİRTAŞ'a sorulabilir.
- Demo sırasında algoritma, geliştirdiğiniz kodun çeşitli kısımlarının ne amaçla yazıldığı ve geliştirme ortamı hakkında sorular sorulabilir.
- Kullandığınız herhangi bir satır kodu açıklamanız istenebilir.