

Yazılım Laboratuvarı 1.Proje Ödevi

1st Umut Şahin Demirbaş
Bilgisayar mühendisliği 2.öğretim
Kocaeli üniversitesi
Kocaeli Türkiye
umutsahin.demirbas37@gmail.com

2nd Emircan aygün
Bilgisayar mühendisliği 2.öğretim
Kocaeli üniversitesi
Kocaeli Türkiye
emir616116@gmail.com

Abstract—Bu döküman kocaeli üniversitesi bilgisayar mühendisliği bölümü yazılım laboratuvarı dersi 1.proje ödevi için hazırlanmıştır

Index Terms—Web scraping,E-ticaret

I. ÖZET

Bu projede webscraping yaparak 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.

II. GİRİŞ

Projede amaçlananlar şunlardır: Web scraping ile farklı alışveriş sitelerinden istenilen ürün hakkında bilgi elde etmek. Bir uygulama içerisinde istenilen özellikteki ürünlerin filtrelenmesi ve sıralanması becerisini geliştirmek . Dinamik özelliklere sahip bir program geliştirmek Web kodlama hakkında bilgi ve beceriye sahip olmak. Projede yapılacak olan web scraping işleminde amaçlanan şey şudur: 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 kendimiz bir web sitesi oluşturacak (sadece bilgisayar(notebook) kategorisi olacak) ve web scraping kullanarak sitedeki bilgiler alınacaktır.Web Scraping kullanılırken siteye ait html bilgiler veya siteye request isteği ile bilgiler alınmalıdır. Veritabanı kısmında ise Web scraping ile elde edilen veriler veritabanına kaydedilmelidir .Kullanılacak veritabanı proje uygun olacak şekilde seçilmelidir Veritabanındaki tablolar normalizasyon kurallarına uygun olacak şekilde tasarlanmalı ve organize edilmelidir. Web Sayfası kısmında istenenler şunlardır Farklı sitelerden çekilen ürün bilgilerinin gösterilmesi için bir web sayfası oluşturmamız. Bu sayfa ilk açıldığında veritabanında yer alan tüm kayıtlar listelenmelidir. Listeleme işleminde ürün bilgilerinin yer aldığı ana başlık bulunmalıdır. Herhangi bir ürün başlığına tıklandığında ürün hakkındaki bilgilerin (fiyat ve ürün özellikleri) olduğu sayfaya yönlendirilmelidir 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. 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 yanlış olması durumunda sistem düzeltilmiş öneride bulunmalıdır. Örnek: hepsi burada – yazımı düzeltilmiş sonuçları görüyorsunuz: Hepsiburada şeklinde olmalıdır. Web sayfasında dinamik filtreleme işlemi de ayrıca yer almalıdır. Filtreleme işlemi veritabanında yer alan ürünün tüm özellikleri için uygulanabilir olmalıdır. 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 olmalıdır. Sitede duplicate kontrolü de yapılmalıdır duplicate kontrolünün amacı ve olma nedenleri şunlardır : Scraping ile çekilen bilgiler güncellenmesi gerekmektedir. Güncel bilgiler veritabanında kaydedilmelidir 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. 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 diğeri unique bir etiket altında temsil edilmelidir. Projede bizden istenenler bunlardır.

III. YÖNTEM

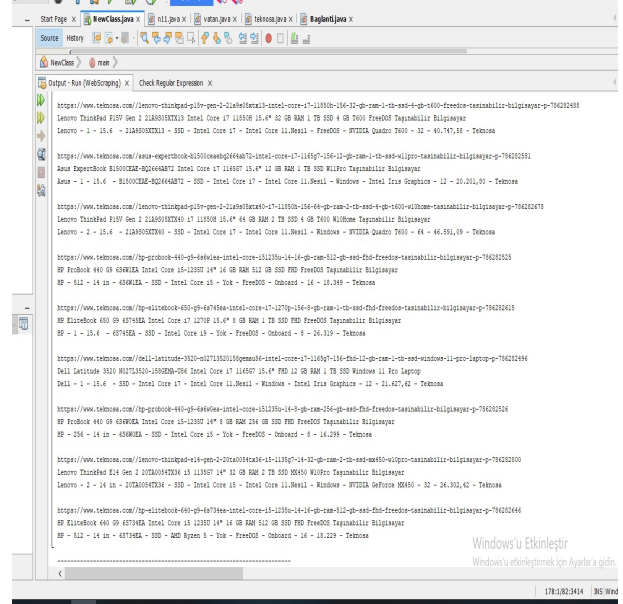
Proje başlangıcında ilk olarak java dili kullanarak web scraping yapacağımızdan gerekli bir araç olan jsoup indirme ve kurulumunu tamamladık.Ardından jsoup aracını, kullandığımız netbeans idesine tanıttık.Gerekli kütüphane importlarını da yaptıktan sonra web scraping işlemine geçtik.İlk aşamada ,önceden belirlemiş olduğumuz 3 site olan vatan, n11 ve teknosa için ayrı bir class tanımlayarak ürün linklerini çekme işlemini bu classlarda yaptık.Çektiğimiz linkleri bir java listesine attık ve bunu ana classımız içinde kullandık.Ana classımız içerisinde gerekli tanımlamaları yaptıktan sonra ilgili

sitelerden gerekli özellikleri çekmeye başladık.Bu ana class içerisinde 4 adet jsoup importu ve 3 adet de sql importu kullandık.if else blokları yarımıyla her bir ürünün çekeceğimiz özelliğini kontrol ederek anlık sonuçları görmek adına ekrana yazdırdık.Çıktılarda doğru sonuçların yakalandığını gördükten sonra ise bağlantı classımız ve jsoup aracımız sayesinde her bir özelliği ,mysqlde yapmış olduğumuz databasedeki tablonun ilgili kolonuna yazdırdık.ve bu işlemleri ana class içerisinde 3 e ticaret sitesi için de gerçekleştirdik. Site yapımı kısmında öncelikle php kodları üzerinde çalışabilmek için Visual studio code uygulamasını indirdik daha sonrasında lokalde php kodlarını görebilmek için lokal server uygulaması olarak XAMPP 1 indirdik çalışma uygulamalarımızı indirip kurduktan sonra projemizde kullanmak için bir tema indirdik bu temadan önce işimize yarayan kısımları aldık sonrasında temayı parçalayıp kendi sitemize uygun hale getirdik sonrasında visual studio code uygulamasından html kodlarımızı açtık ve düzenlemeye başladık düzenledikten sonra .php uzantısıyla dosyalarımızı kaydettik siteye eklemek üzere bir logo tasarladık bunu ekledik daha sonrasında site sayfalarında kullanmak üzere bir database oluşturduk bu database ile sitemizi bağladık admin paneli ve ürünler sayfasını oluşturduk database e eklediğimiz deneme amaçlı ürünler ile sitenin çalışıp çalışmadığını kontrol ettik.Sonrasında sitemiz için php kodlarında kategoriler ve arama kısmını yaptık fiyat aralığını ayarlayabileceğimiz bir buton ekledik. Admin panelinden ürün ve kategori ekleme sistemini yaptık.Database kısmını oluştururken Phpmyadmin i admini kullandık phpmyadminden tabloları oluşturduk ve sitemize örnek olması açısından bir kaç adet ürün ekleyip siteyi kontrol ettik.

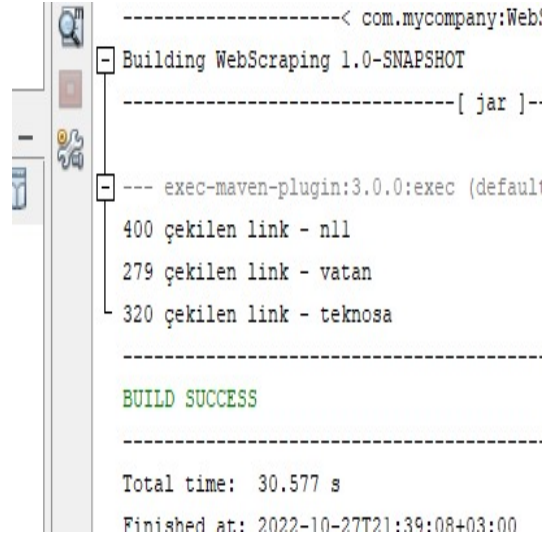
IV. SONUÇ

Sonuç olarak bu projeden websitelerinden bilgi çekmeyi yani Web scraping işlemini deneyimlemiş olduk.İstediğimiz siteden istediğimiz bilgileri nasıl çekebileceğimizi öğrendik.Bunun dışında internet sitesi kodlamada yani web kodlamada bilgi edindik yeni bir yazılım dili olarak PHP ve html hakkında bilgi edindik site yapmanın database oluşturmanın zorluklarını gördük ve deneyimledik.Her gün girdiğimiz nasıl yapıldığını çokta önemsemediğimiz o sitelerin altında yatan mimari hakkında bilgi edindik.Lokalde bir web sayfasının nasıl açılacağını öğrendik XAMPP uygulaması sayesinde lokal bir sunucu oluşturmayı ve bu sunucuda phpmyadmin ile database oluşturmayı bu database ile etkileşime girmeyi öğrendik. Php kodlarının lokalde çalışmadığını ancak bir server aracılığı ile çalışabileceğini öğrendik.

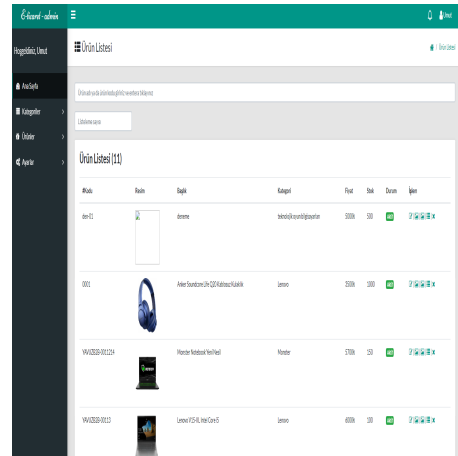
V. DENEYSEL SONUÇLAR



Web scraping sonrası çekilen verilerin ekran görüntüsü



Web scraping sonrası çekilen verilerin sayısı



site admin paneli

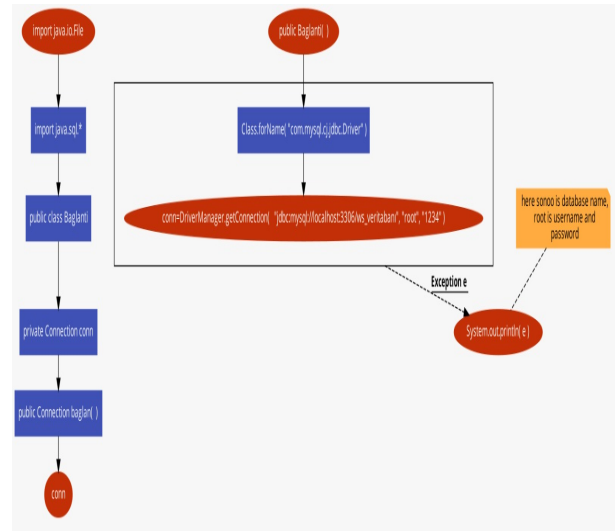
site anasayfası

site ürün sayfası

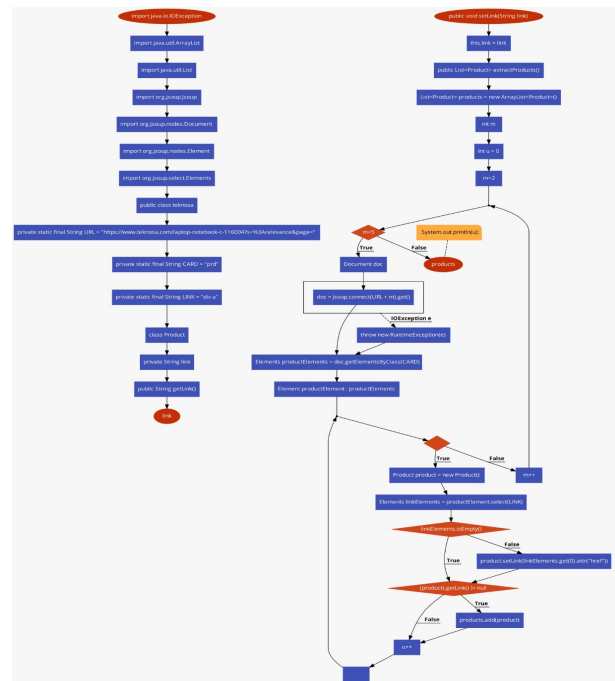
REFERENCES

- [1] <https://themeforest.net/item/hurst-e-commerce-furniture-template/16921717>
- [2] <https://www.youtube.com/watch?v=mg-588VDyuIt=0s>
- [3] <https://www.youtube.com/watch?v=H2yeLbzes0>
- [4] <https://www.webscrapingapi.com/java-web-scraping>
- [5] <https://www.scrapingbee.com/blog/introduction-to-web-scraping-with-java/>

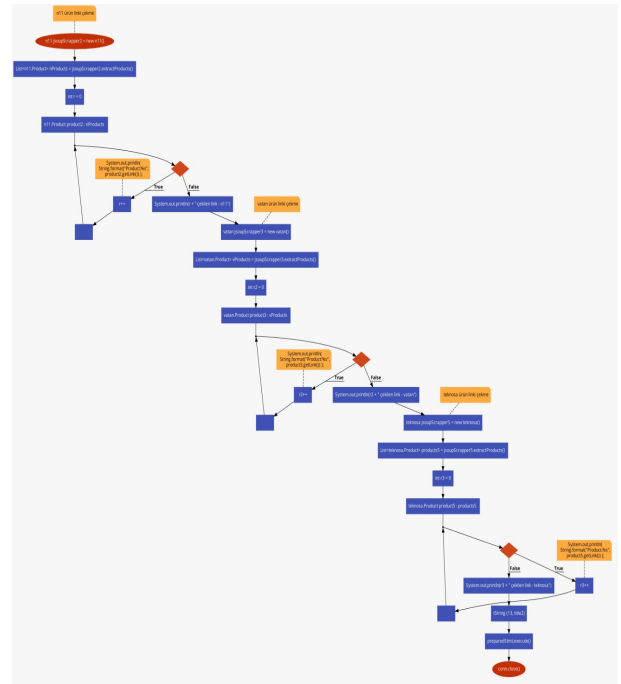
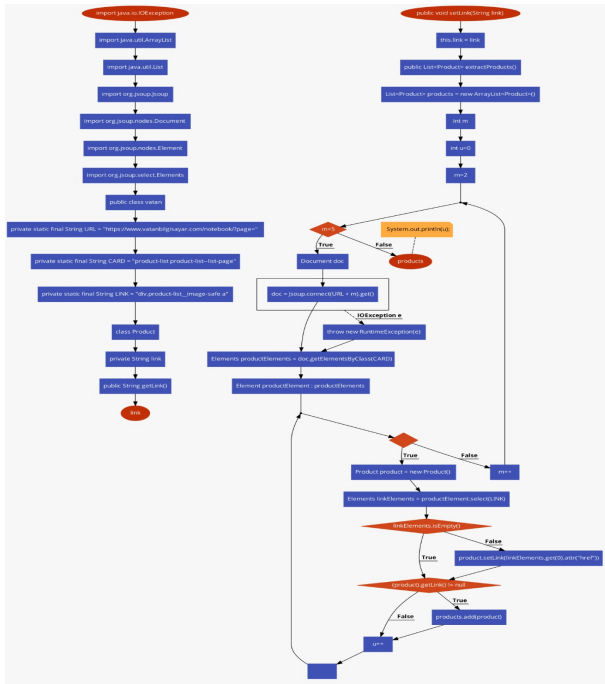
VI. AKIŞ DİYAGRAMLARI



j soup bağlantı classı

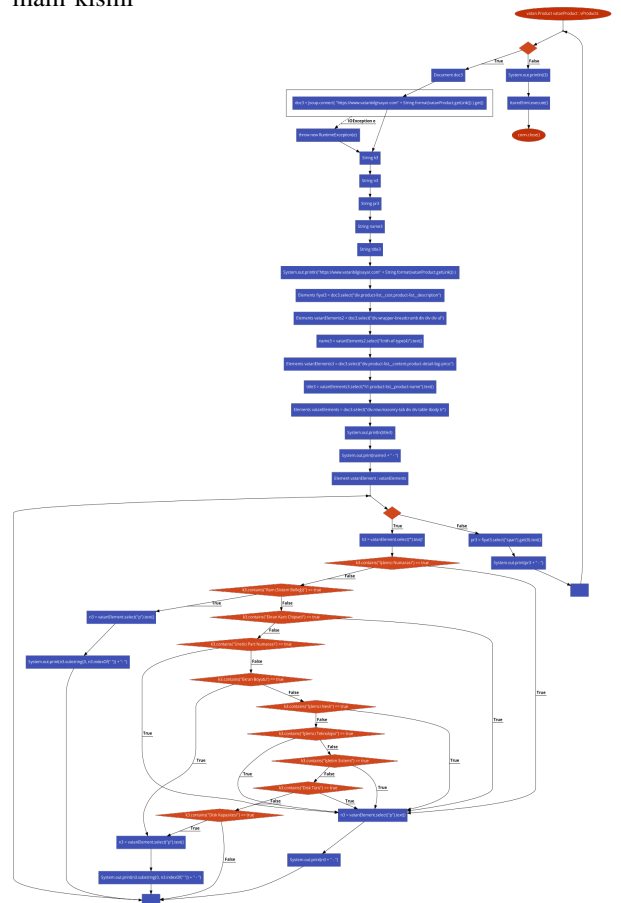
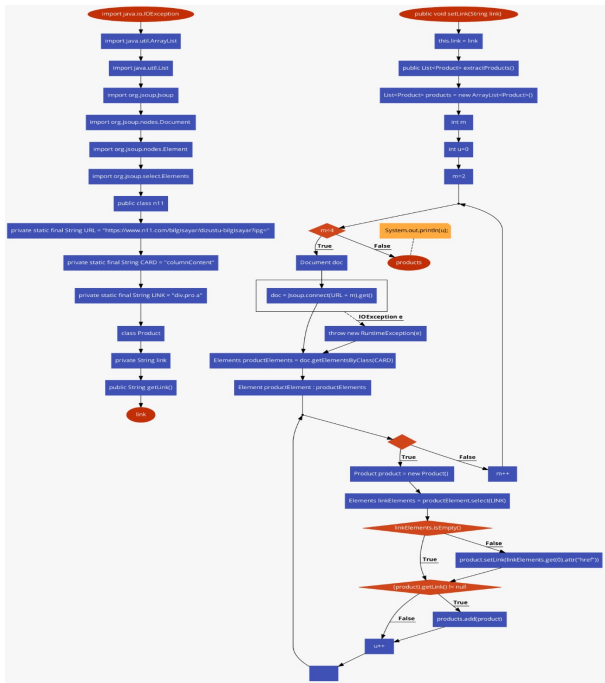


Teknosadan ürün linkleri çektiğimiz class



main kısmı

Vatandan ürün linkleri çektiğimiz class



Vatan ürün özellikleri çekme kısmı

N11den ürün linkleri çektiğimiz class