

## Modül 1 - Day 2

### Modül-1: Data Collection

**İçerik:** API'lar, loglama, sensory data, web scraping.

**Anahtar sözcükler:** JSON, XML, HTTP, HTML, DOM, grep, RegExp.

**Araçlar:** Postman, log4j, python-logging, BeautifulSoup, Jsoup, Selenium

**Postman**, 3 milyondan fazla yazılımcının API'ları paylaşmak, test etmek, dokümanete etmek, monitör etmek için kullandığı, dünyanın en çok kullanılan "rest client"ı. En öne çıkan özelliği tüm bunlar için çok kullanışlı bir arayüz sunmasıdır.

### HTTP

"HTTP" bilginin sunucudan kullanıcıya nasıl ve ne şekilde aktarılacağını gösteren protokoldür. İnternet ağında sunucular ve kullanıcılar arasında nasıl bir veri alışverişi olacağı hakkında kurallar vardır. Bu düzeni sağlayan protokol de HTTP'dir İnternet sitesine girmek için adres çubuğuna sitenin adresini yazdığınız vakit HTTP ile sunucuya bir istek gönderilir ve sunucu bu isteğe cevap verdiği vakit internet sitesinin verileri size gelir.

### Restful Mimarisi

RESTful servisler veri iletiminde farklı HTTP metotlarını kullanmaktadır. Yapılan HTTP request'i için çağrılan URL ile beraber HTTP method bilgisi bahsi geçen 4 metottan biri olarak seçilir ve sunucu yapılan talebin kayıt üzerine nasıl etki edeceğini buna göre belirler.

**Ç ok sık kullanılan iki fonksiyonu bulunur:**

- GET: Veri listeleme - veri görüntülemek için kullanılır. Örneğin arama motoruna trednyol.com yazıp enter' a basarsam http get isteği yapılmış olur. Basit bir şekilde gizlemeye gerek kalmadan bilgiyi açık şekilde gönderir. Format eklenebilir.
- POST: Veri eklemek için kullanılır. Trendyol sitesine giriş yapma, kredi kartı bilgileri girme vs işlemler yani aktivite gerektiren olaylar posttur.

**NOT:** Get bir nevi read işlemiyken post write işlemine benzetilebilir. Server'a bilgi gönderilir. Post da gönderilen bilgiler gizli şekilde post datası olarak gönderilir.

### Diğer Metodlar

- PUT: Veriyi Güncelleme isteği olarak kullanılır. Bunun yerine post metodu da kullanılabilir.
- PATCH: Verinin sadece bir parçasını güncellemek için kullanılır. Örneğin bir issue'nun durumunun aktiften çözüldü haline getirilmesi.
- DELETE: Veriyi silmek için kullanılır. Bunun yerine post metodu da kullanılabilir
- OPTIONS: Bir api urline Options isteği yapıldığında o url in hangi istekleri kabul ettiği bilgisi verilir.

<http://httpbin.org/> bu adresyen post- get ya da diğer metodların örneklerini deneyebilirsiniz. Bu konuyu ayrı bir yazı olarak paylaşacağım.

## CURL Komutu

Çoğu Unix bazlı sistemde mevcut olan bir komuttur ve “Client URL”nin kısaltılmışıdır. Curl komutları URL’lerin bağlanabilirliğini kontrol etmek ve veri transferi için harika bir araç olarak kullanılmak için üretilmiştir. Curl komutları aşağıdaki protokolleri destekler:

- HTTP ve HTTPS
- FTP ve FTPS
- IMAP ve IMAPS
- POP3 ve POP3S
- SMB ve SMBS
- SFTP
- SCP
- TELNET
- GOPHER
- LDAP ve LDAPS
- SMTP ve SMTPS

Başka desteklenen komutların da bulunmasıyla beraber listedeki komutlar en önemli protokollerdir. Curl, istemci tarafındaki ücretsiz bir URL transfer kütüphanesidir.

```
curl -X GET "http://httpbin.org/get" -H "accept: application/json"
```

Curl komutunun derste kullanıldığı bir örnek şekilde gösterilmiştir.

## URL içerisindeki özel karakterler

- "?": "sorgu parametreleri" olarak adlandırılır. Sunucu, bu ek bilgilere dinamik olarak cevap verebilir ve göreceğiniz sayfayı bu parametrelere göre oluşturabilir. Bu, sayfadaki bir alana otomatik olarak yazılmış bir bilgi veya bir arama motorundaki arama sorgunuz olabilir.
- "%": "Escaping" olarak adlandırılan bu işlem, URL'deki boşluk karakterinin soruna yol açmaması için alternatif bir biçime dönüştürülmesidir. Örneğin +yazdığınızda bu karakter, %3F ile değiştirilir.
- "=": anahtar-değer çiftlerini temsil eder. Anahtar-değer çiftine bir örnek, sayfa=4 olabilir. Burada "sayfa" anahtar, "4" ise değerdir.

## GREP Komutu

Elimizdeki loglarda buradan terminal ekranında hızlıca sorgularımız getirebiliriz. Logların güzel şekilde tutuluyor olması işimizi kolaylaştırıyor. Pipe ile çok daha etkin kullanım yöntemleri mevcuttur.

Ders Örnekleri:

```
grep "GET" * —color
```

```
grep "42.236.10.125" * —color | wc -lw
```

```
grep "mozilla" * —color
```

## En sık karşılaşılan hata kodları

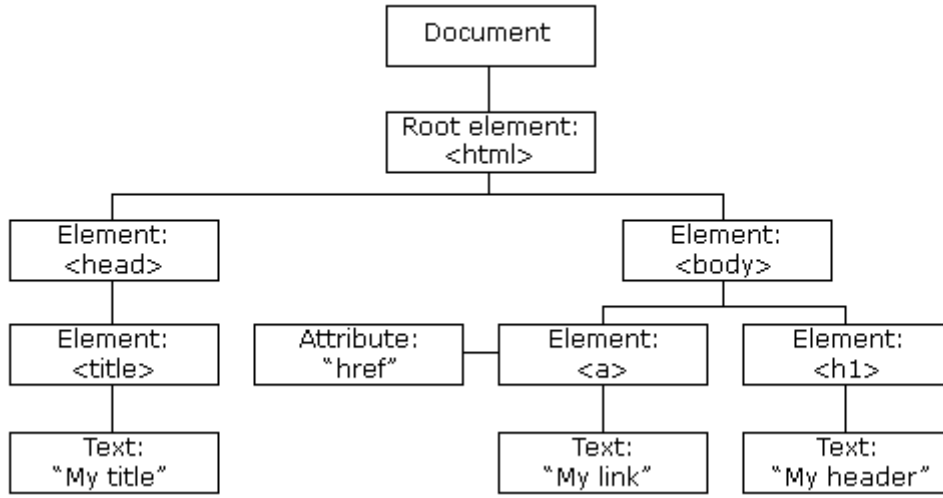
**HTTP 200 (OK):** Yanıtın başarılı olduğunu gösterir. Yani, istemci ile sunucu arasındaki iletişim herhangi bir hata olmadan sorunsuz bir şekilde yürütülmüştür.

**HTTP 404 (Not Found):** İstenen kaynağın sunucu tarafından bulunamayacağı anlamına gelir. Bu, geçici bir aksaklıktan kaynaklanıyor olabilir ve gelecekte başka bir istekte bulunulursa kaynak kullanılabilir olabilir. Çoğunlukla, 404'e götüren bağlantılara genellikle **bozuk bağlantılar** denir.

**HTTP 502 (Bad Gateway):** Sunucunun proxy olarak hareket ederken istekte bulunurken sunucudan geçersiz bir yanıt aldığını gösterir.

### DOM(Document Object Model)

HTML için kullanılan doküman nesne modelidir. HTML Elementlerini objeler olarak, HTML elementlerinin tüm özelliklerini, HTML elementlerine erişmek için metotları, tüm HTML elementleri için olayları tanımlar. Diğer bir deyişle HTML DOM yeni elementler eklemek, elementleri değiştirmek veya silmek için kullanılır



HTML DOM Nesnelerin Ağacı

### BeautifulSoup4

**Kaynak:** <https://medium.com/@nuriyavuz2.71/python-beautifulsoupmodülü-689ef499ee16>

BeautifulSoup, HTML veya XML dosyalarını işlemek için oluşturulmuş güçlü ve hızlı bir kütüphanedir. Bu modül ile bir kaynak içerisindeki HTML kodlarını ayrıştırıp sadece istediğimiz alanları kesen programlar, daha popüler adıyla BOT yazabilirsiniz.

*Ben pip3 ile kurmuştum, pip3 python3 için paket yöneticisi.*

sudo apt-get install python3-pip

*komutu ile pip3 kullanmaya başlayabiliriz. Eğer pip3 yüklüyse buna gerek yok, şimdi bu modülü kuralım.*

pip3 install beautifulsoup4

### Selenium

**Kaynak:** <https://www.sinanerdinc.com/python-selenium-modulu-kullanimi-1>

Selenium, bilgisayarınıza yükleyeceğiniz bir driver yardımı ile ekrana chrome, firefox gibi bir tarayıcı açarak, gerçek bir insan gibi istediğiniz tüm işlemleri programlama dili yardımıyla çalıştırmanızı sağlayan bir araçtır.

GISTLER:

BS4: <https://gist.github.com/sirin/695abacaa207ad7af20f18c946d19858>

Selenium: <https://gist.github.com/sirin/0e1491163b8f485a476e0991ad228b86>

İkinci gün sonu 😊