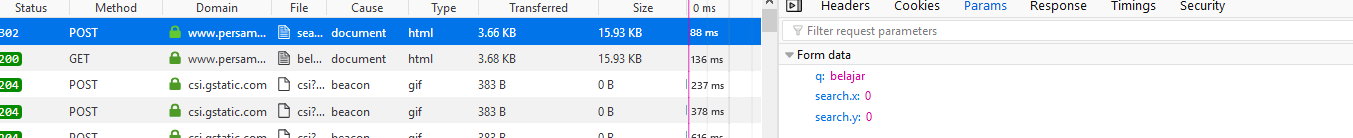
**Simple Web Scraping with Python**

Ada 2 cara untuk mengekstrak data dari sebuah website, yaitu:

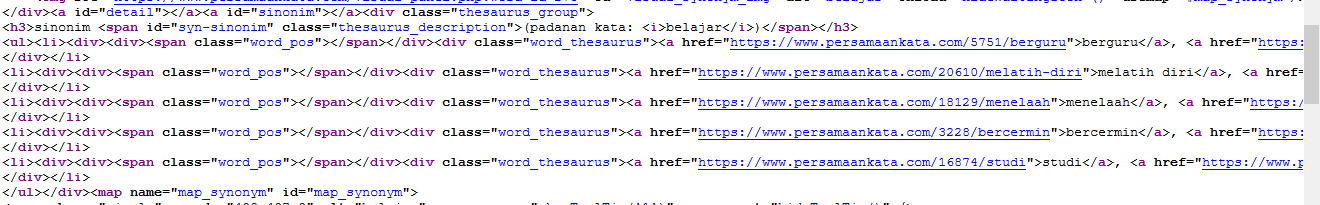
1. Dari API website tersebut(jika ada)
2. Dengan mengakses HTML sebuah website dan mengekstrak data untuk mendapatkan informasi/data

Pada sharing kali ini kita akan melakukan web scraping dengan python.

* Pertama kita akses web yang akan ambil data nya. Disini kita targetkan <https://www.persamaankata.com> . karena kita akan mengambil data hasil dari cari persamaan kata, maka kita coba untuk mencari suatu kata, sambil melihat aktifitas network pada browser biasanya bisa diaktifkan dengan F12. bisa dilihat method yang digunakan saat mencari kata ada lah POST dan params yang digunakan adalah ‘q = kata’. Itu yang akan kita gunakan.



* Setelah itu kita lihat juga source dari HTML yang tampil, pada browser bisa klik kanan > ‘view page source’. Kemudian dari cari data yang akan kita ambil, disini yang akan kita ambil adalah hasil pencarian persamaan kata, seperti dibawah ini:



Dapat dilihat disana data yang kita targetkan ada dalam tag ‘div’ dengan class ‘thesaurus\_group’ dan berada dalam teks dari ‘href’ (link).

* Disini kita menggunakan library Requests untuk mengirim request dan menerima data dari web yang ditargetkan. BeautifulSoup dan html5lib untuk mengekstrak data dari HTML.
* Karena method yang digunakan POST maka penggunaan requests akan seperti ini:

res = requests.post(url,data=params)

* Lalu untuk penggunakan BeautifulSoup dan html5lib seperti ini:

data = bSoup(res.content,'html5lib')

* Untuk mengambil data pada tag ‘div’, gunakan ‘find’

temp = data.find('div',{'class':'thesaurus\_group'})

* Setelah itu untuk mengambil semua ‘href’(link) yang ada dalam div tadi gunakan perulangan:

for i in temp.findAll("a"):

teks = i.text

print(teks)

Untuk source code nya bisa lihat pada example.py