

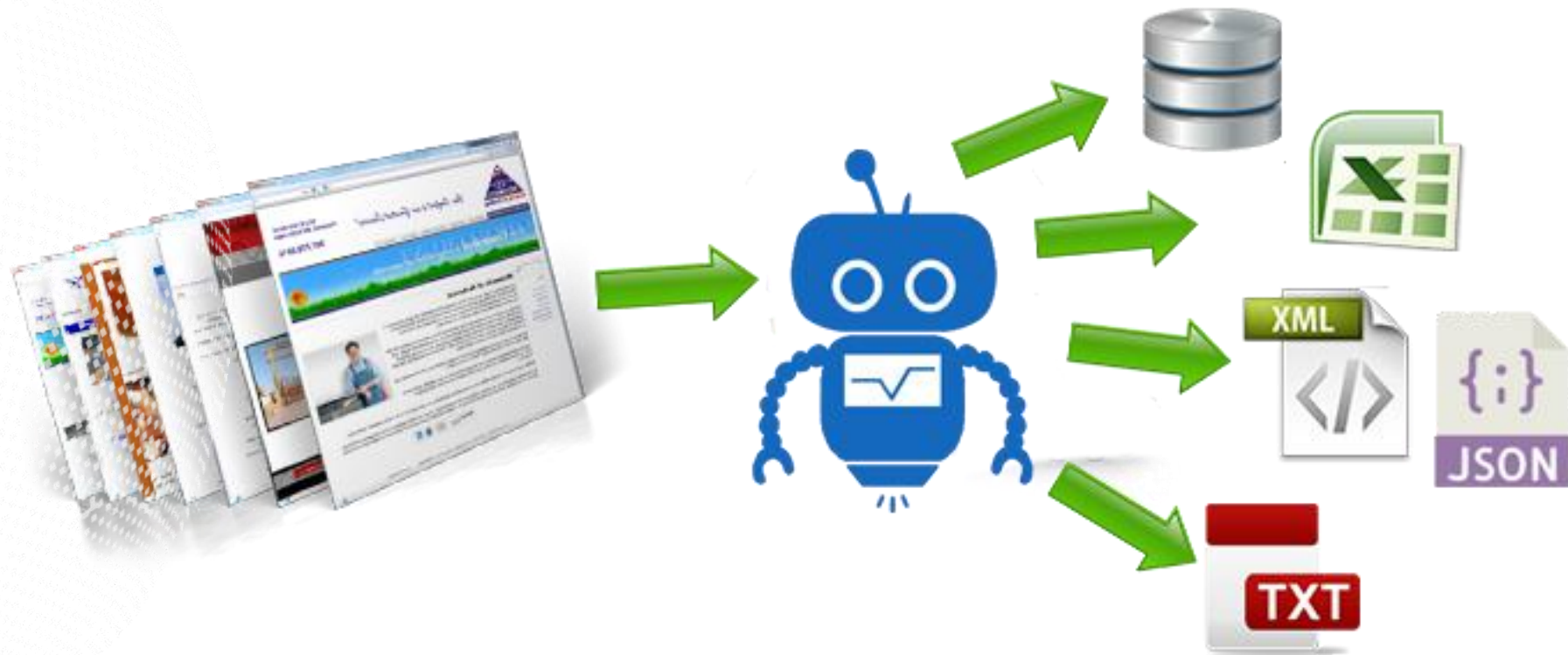


Web scraping

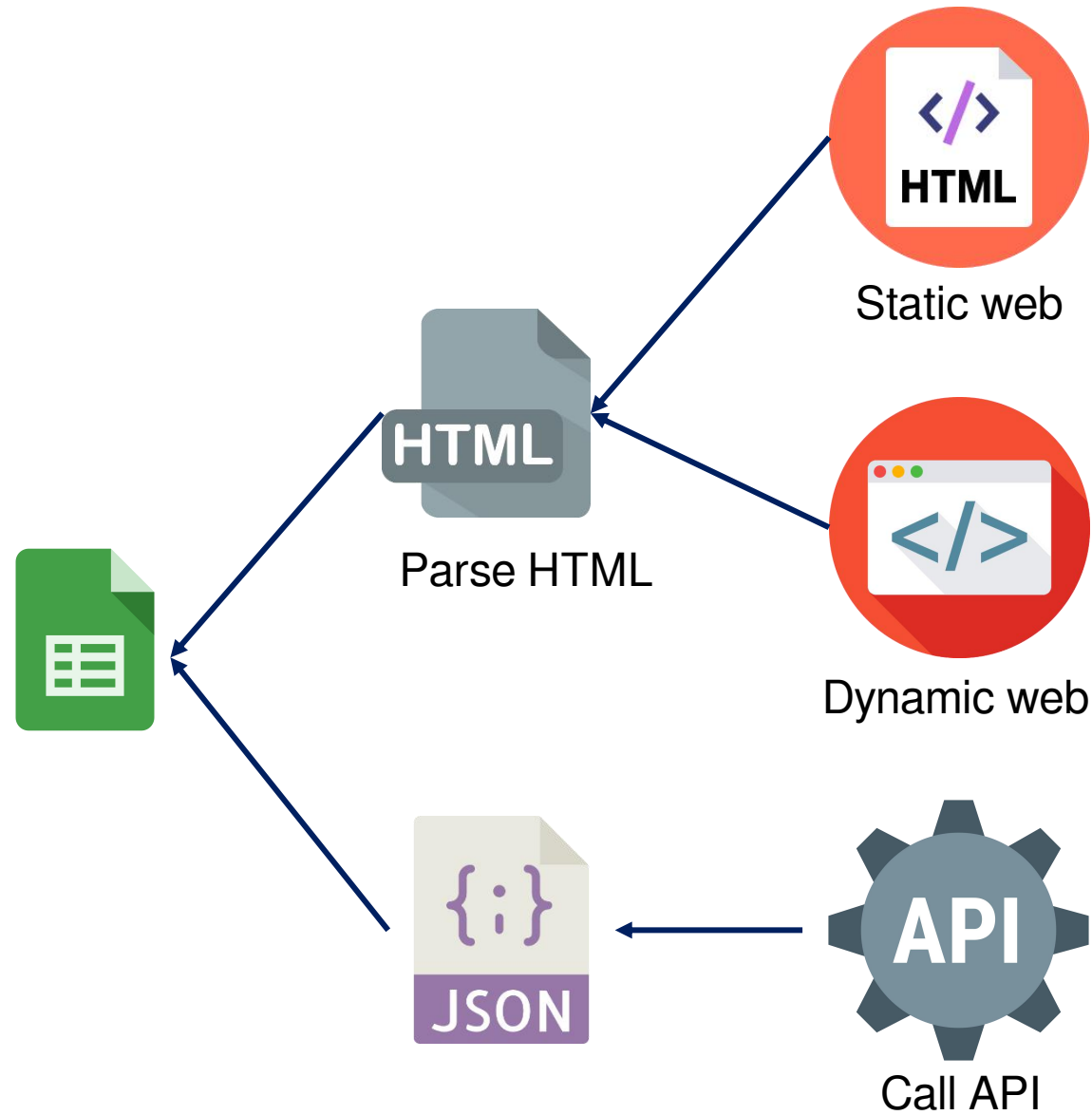
Data scrapping Why?



Web scraping What?

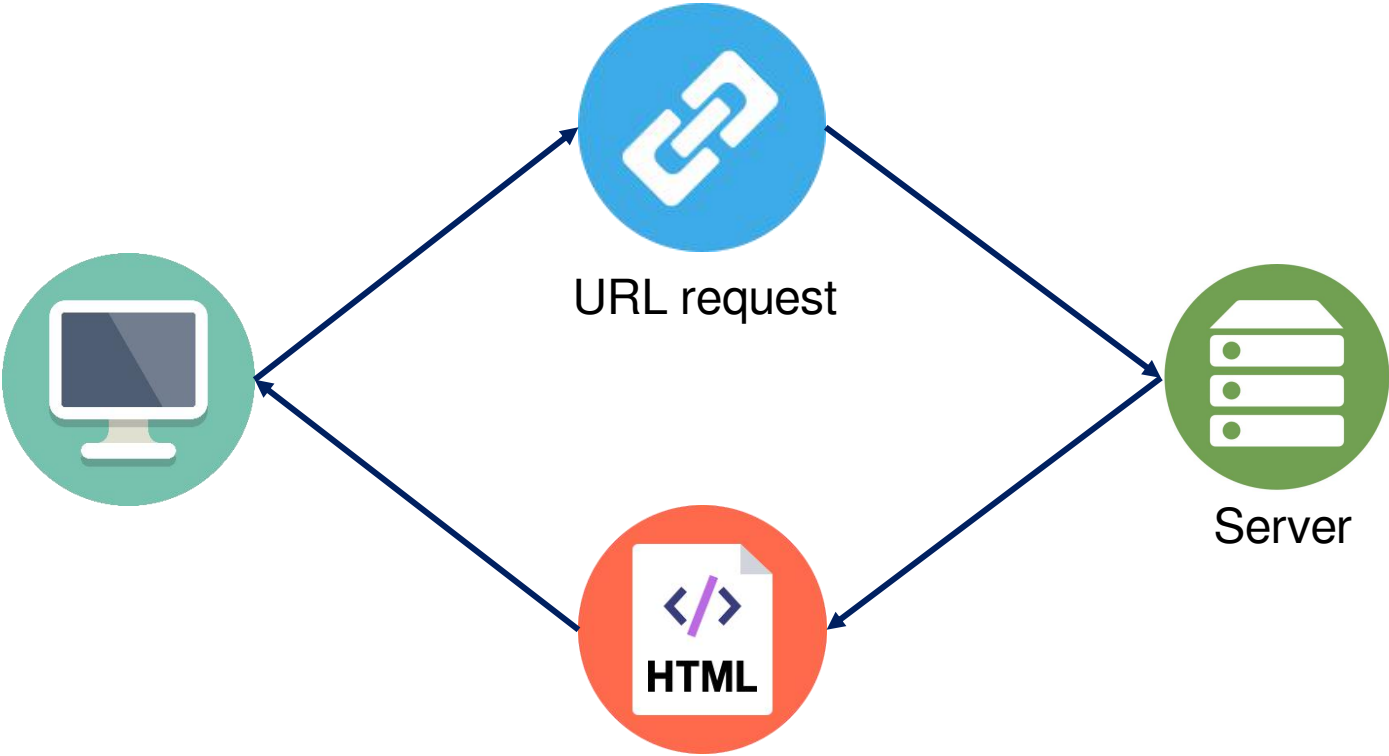


Web scrapping What?

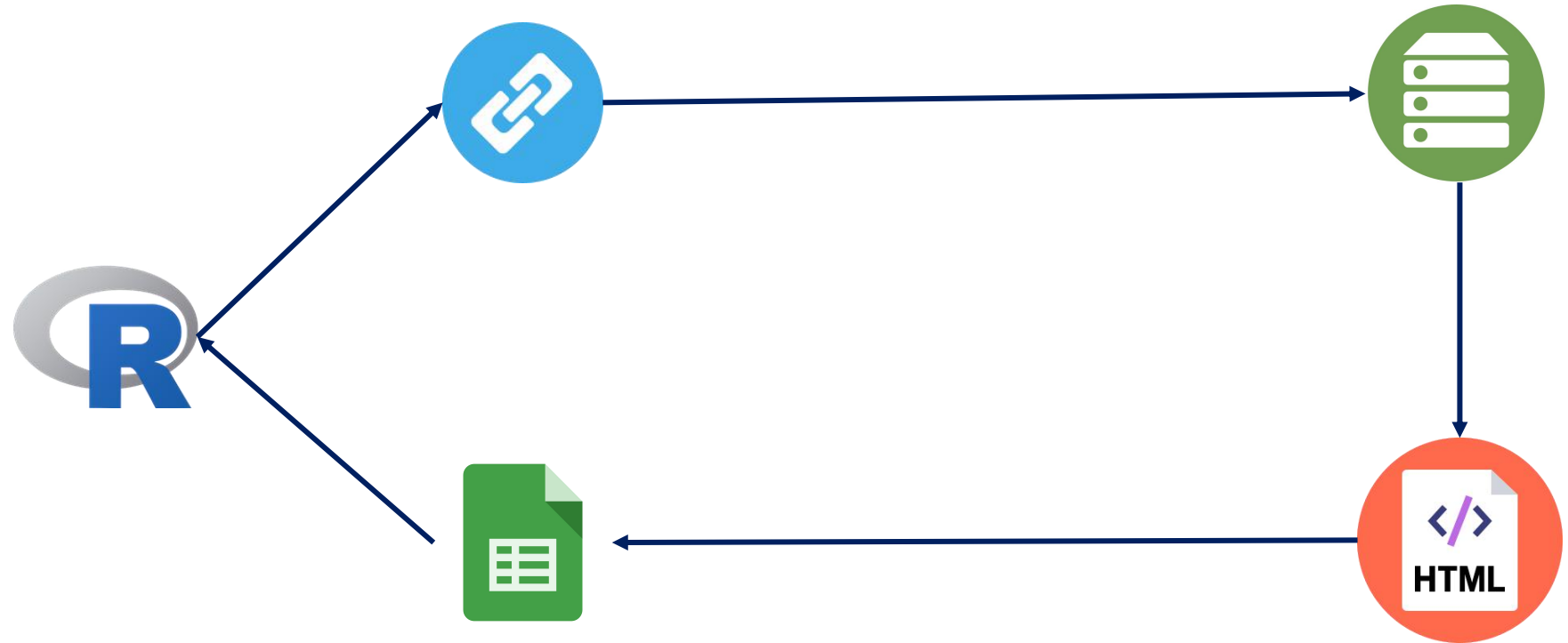




Static web What?



diemthi.tuyensinh247.com/diem-chuan/dai-hoc-tay-bac-TTB.html?y=2016



```
htm %>%  
html_node(xpath= '//*[@id="reBench"]/div[1]/div[3]/table[1]') %>%  
html_table(header = T)
```

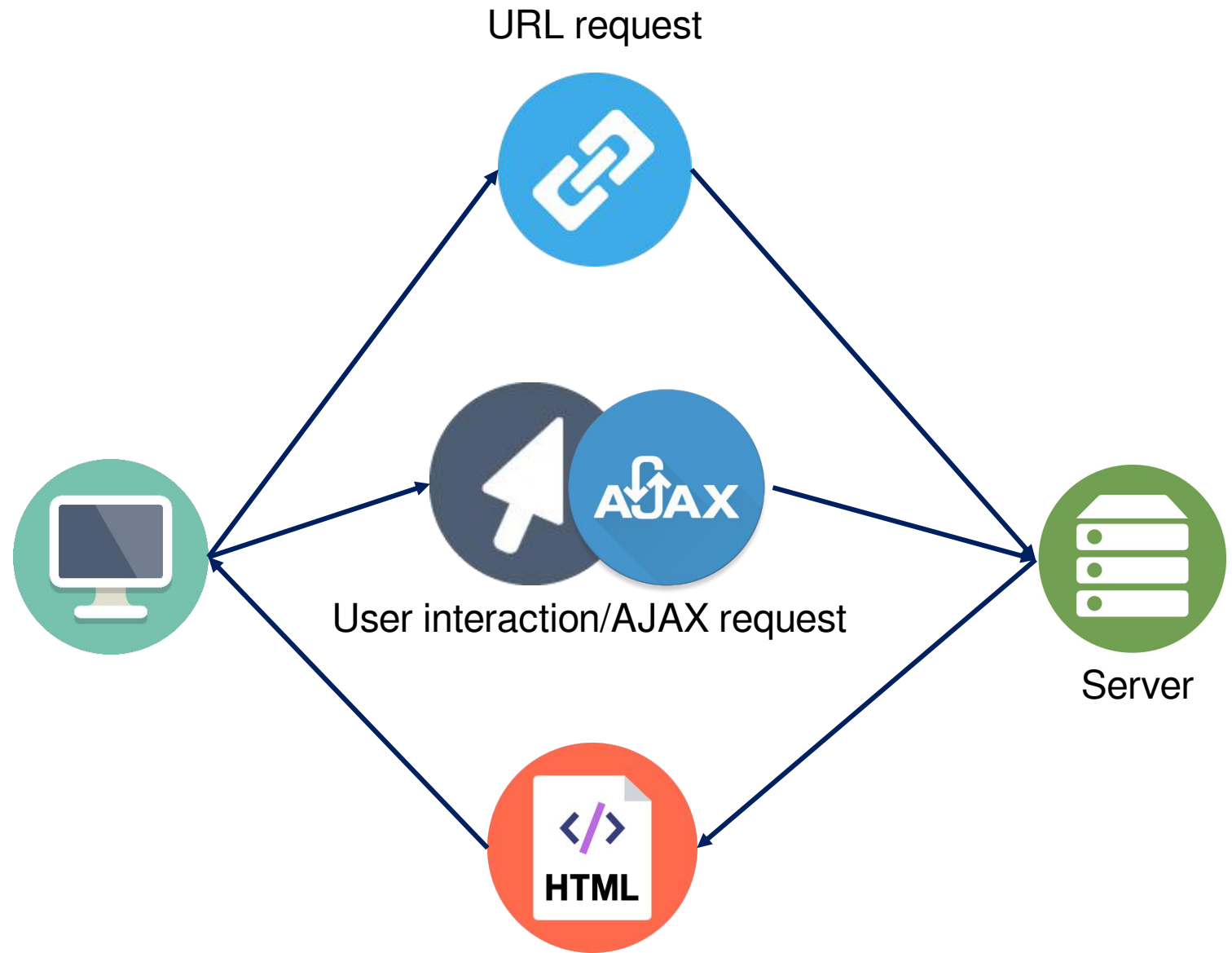
Trường: **Đại Học Tây Bắc - 2016**

Năm: 2016

STT	Mã ngành	Tên ngành	Tổ hợp môn	Điểm chuẩn	Ghi chú
1	51140202	Giáo dục Tiểu học.	A00; A01; C00; D01	11.5	
2	51140201	Giáo dục Mầm non.	M01; M02; M03	11.5	
3	0	Các ngành đào tạo cao đẳng	0	---	
4	52850101	Quản lý tài nguyên và môi trường	A00; A01; A02; B00	15	

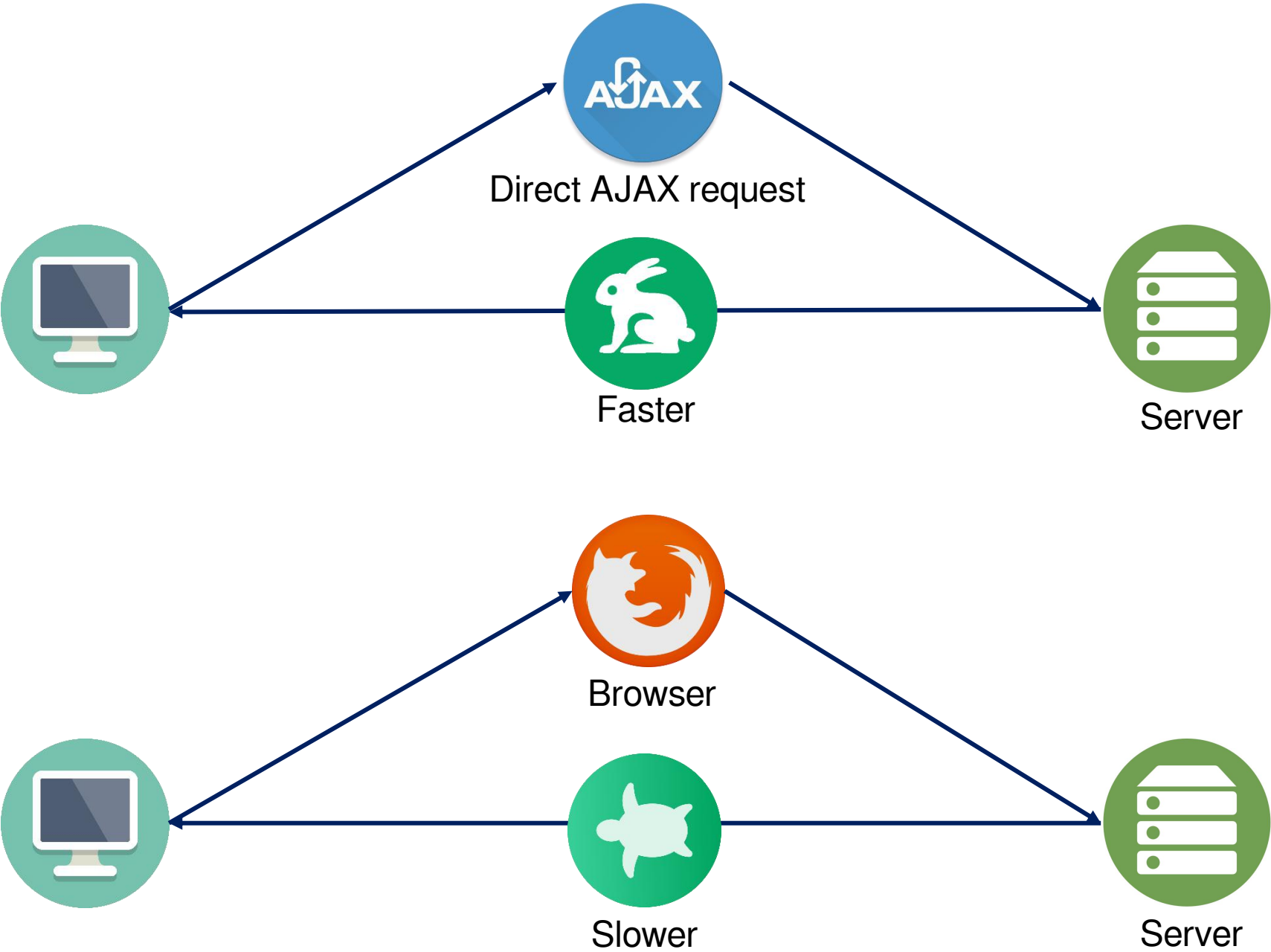


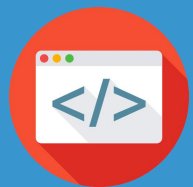
Dynamic web What?



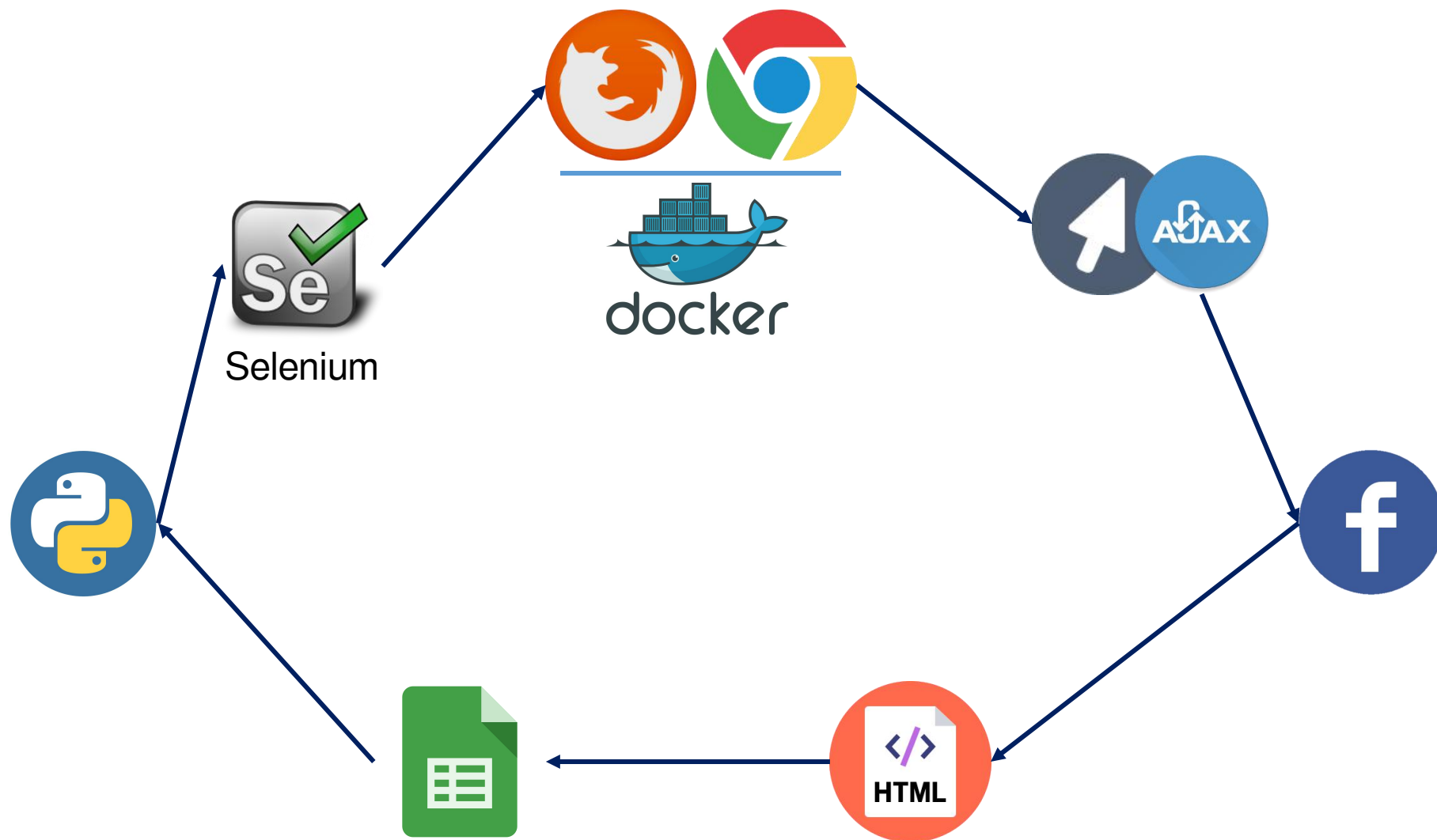


Dynamic web How?





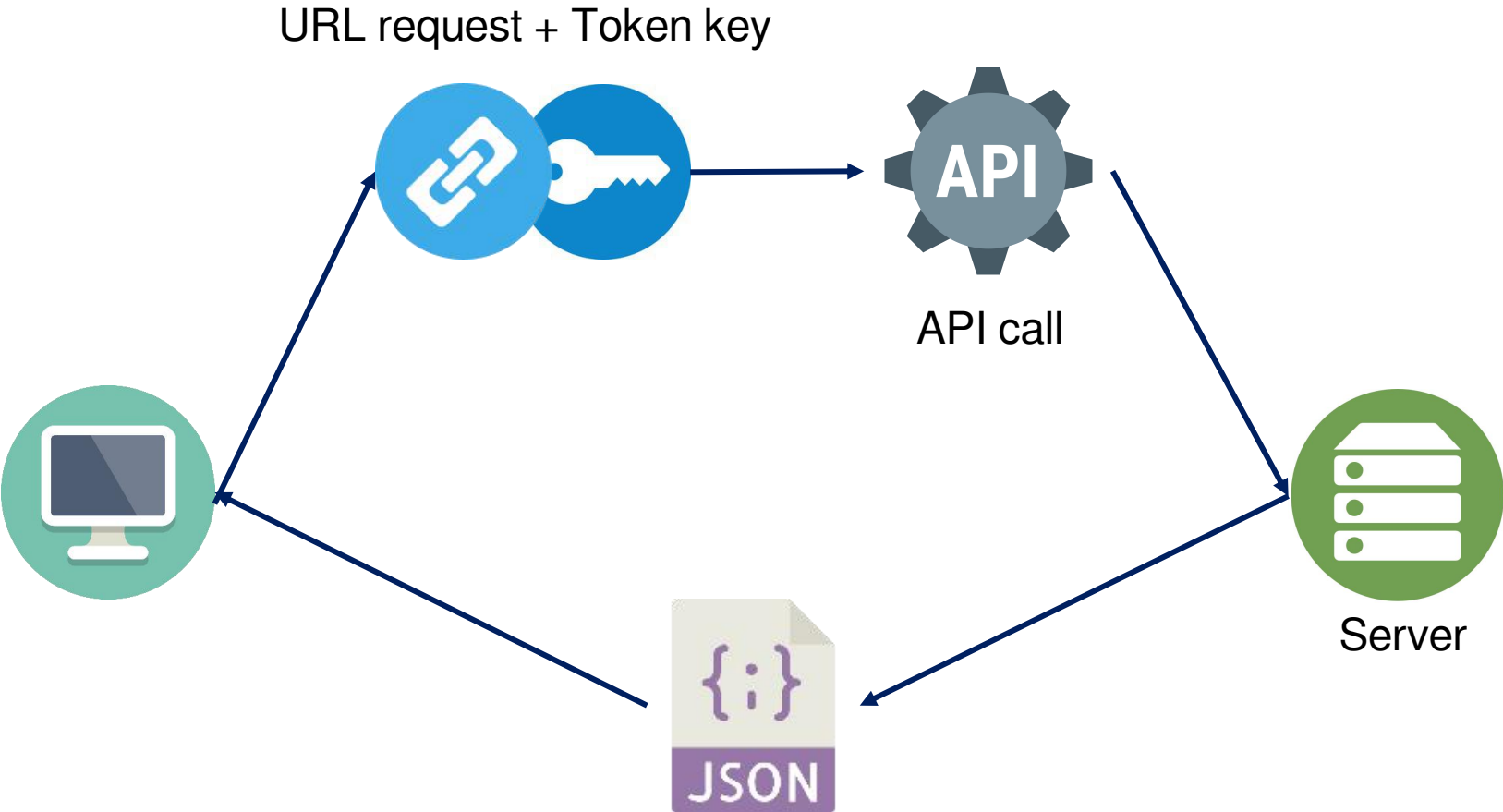
Dynamic web How?



```
driver.  
find_element_by_css_selector('#work > div:nth-child(2)').  
find_elements_by_css_selector('div[id="u_"]')
```

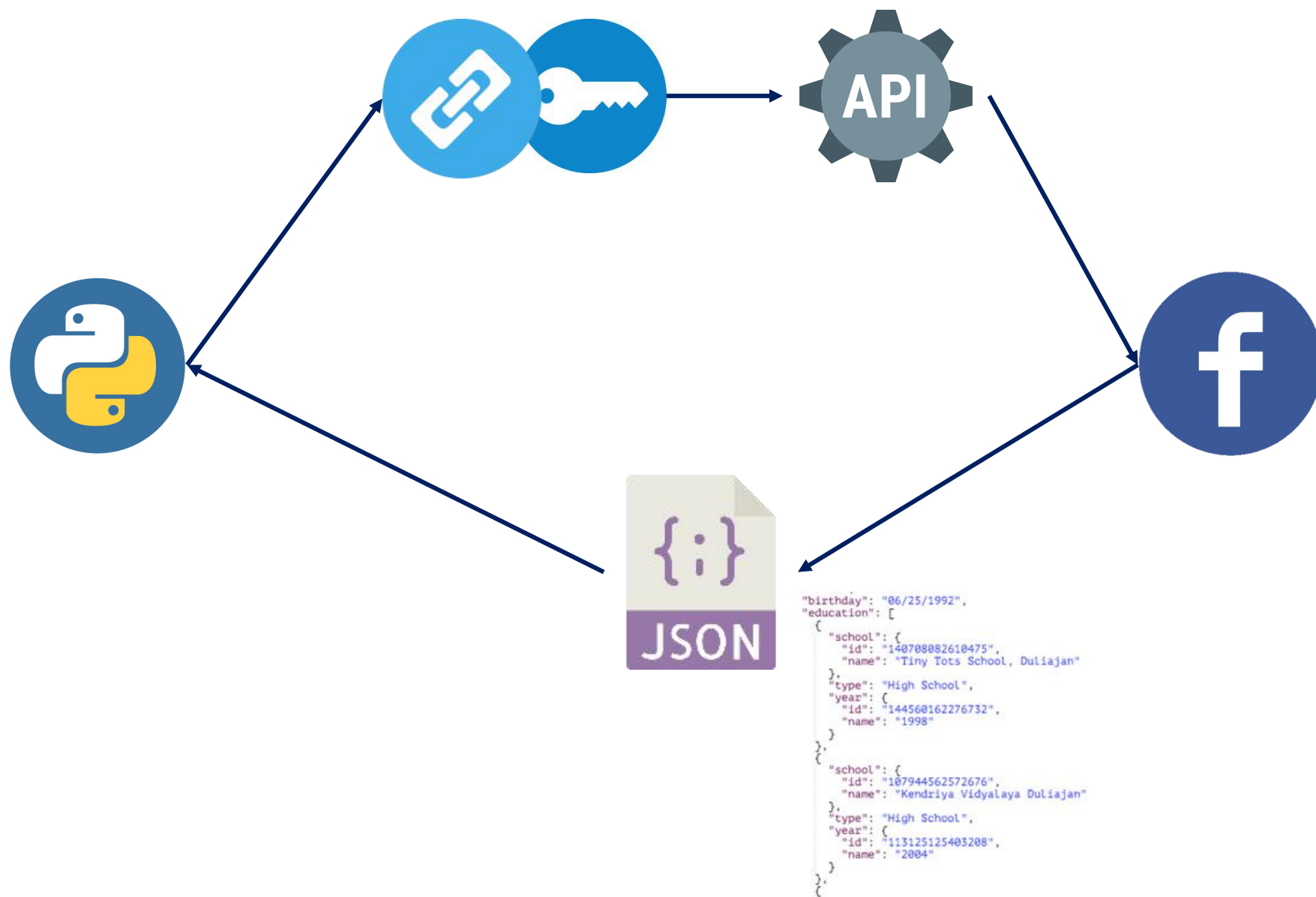


API What?



API How?

```
string = 'https://graph.facebook.com/v2.10/2313165?fields=reactions&access_token=' + 'awcdgweh123123'  
requests.get(string).json()
```



Web scraping Problem

Facebook's anti scraping mechanism

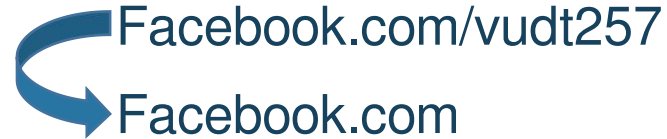
Blank result



We couldn't find anything for **13132661**

Looking for people or posts? Try entering a name, location or different words.

Redirect



Block account

For security reasons your account is temporarily locked

If this account reflects your real name and personal information, please help us verify it.

Continue

Block IP



Network Timeout

The server at _____ is taking too long to respond.

Web scraping Solution



Delay



Switch account



VPN