Anyone can be

Mark Zuckerberg

Shubham

bhawsinka.com @shubhamagr

Yatish

yatishmehta.com @yatish27

Warmup Python Quiz

```
if <BLANK>
    print("I am in if statement")
else:
    print("I am in else statement")
Output:
I am in if statement
I am in else statement
```

Warmup Python Quiz

```
if print("I am in if statement"):
    print("I am in if statement")
else:
    print("I am in else statement")
Output:
I am in if statement
I am in else statement
```

Web Scraping

What is Web Scraping?

- An automated method of browsing and collecting the unstructured web data into structured data.
- Also termed as Web Harvesting
- In 2013, 23% of all web traffic was scraping related*

Why do we need it?

- Web World is a wealthy world

- No API access

- Automation

- Analytics

How do we do it?

- Find the web pages
- Parse HTML

- XPath or CSS selectors using Web inspector, Firebug
- Scrap
- Export data to CSV and Enjoy!

Demo

Web Scraping in Ruby

Why Ruby?

- Beautiful Language
- Purely OO, Ease of coding
- Great community support

- Gems

Nokogiri

- Html/XML Parser

- Easy navigation

- Helper methods

Mechanize

- Full fledged gem for building web bots

- Uses nokogiri

- Similar to beautifulSoup in python but better personal bias;)

Waitr

- Browser Simulation

- Includes javascript engine

- Slower and resource intensive

Demo 2

- Linkedin Scraper, Open source RubyGem, >14k Downloads

- github.com/yatish27/linkedin-scraper

Challenges and Issues

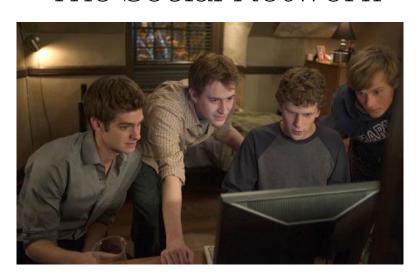
- Improper HTML Markup

- Overloads the web server with requests
- Blockers(robots.txt)

- Legal

Web Scraping Products

First 10 minutes of "The Social Network"



Web Scraping Products

Google !!!



Web Scraping Ideas

Builtwith.com Address/Email/Phone Parsing Search Engines Facebook Analysis Ecommerce monitoring Lead Gen

Questions