Scraping Dynamic Web Pages with Python and Selenium

EXPLORING SELENIUM WITH PYTHON
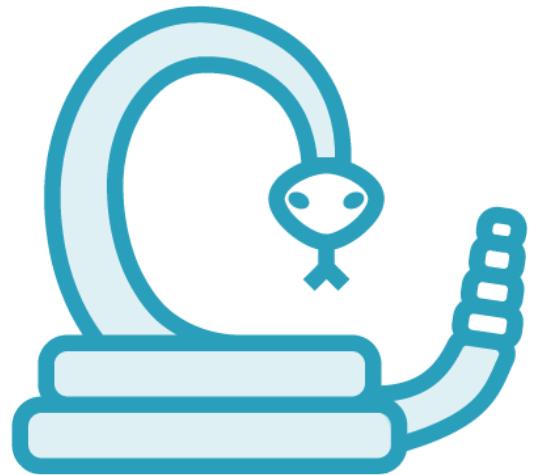
Pratheerth Padman
FREELANCE DATA SCIENTIST

Prerequisites
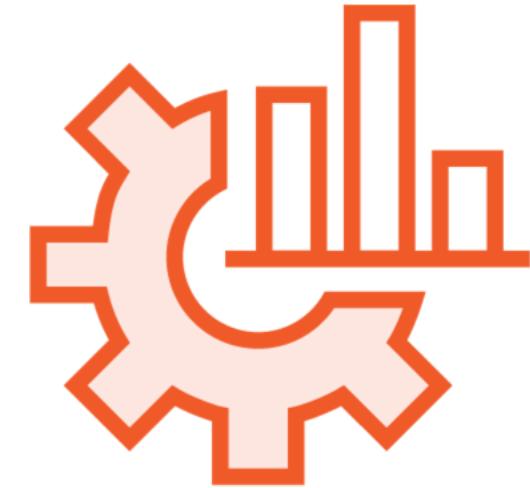
# Prerequisites

**Python programming**
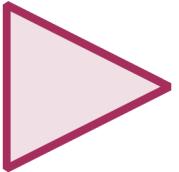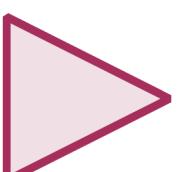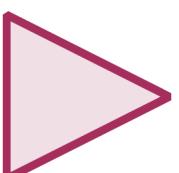
HTML and CSS

Data analytics literacy

Course Pattern

Concept introduction

Concept detailing

Practical implementation

Module Overview

What does scraping dynamic webpages mean?

What is Selenium? Where is it used?

Selenium WebDriver

Demo: Installing required libraries and packages

Demo: Opening a webpage with WebDriver

Overview of Scraping Dynamic Webpages

Web Scraping

Web scraping is the process of extracting information and data from a webpage into a structured format for further analysis.

Types of Webpage Content

Static content

Fixed content. Incapable of action or change.

Dynamic content

Contains dynamically-rendered components and information that changes based on many factors.

What Is Selenium and Where Is It Used?

What is Selenium?

Selenium is a set of tools and libraries that automate web browser actions

Web-driver package to control browser and emulate user oriented actions

Open source

Python, Java, Ruby, Scala, C# etc.

Components of Selenium

Selenium IDE

Used for prototype testing

Selenium RC

**First tool of the selenium suite**

Selenium WebDriver
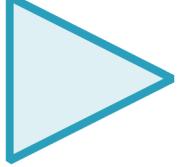
Developed to overcome limitations of RC

Selenium Grid
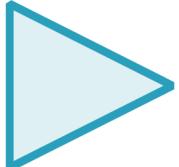
Used for parallel testing
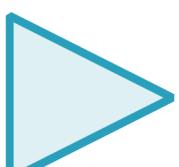
Where Is It Used?

Scripts used to automate any manual user interaction

Automated web application testing

Scraping dynamic web pages

Selenium WebDriver

Selenium WebDriver

Biggest recent change – inclusion of WebDriver API

**Drives a browser natively, as a user would**
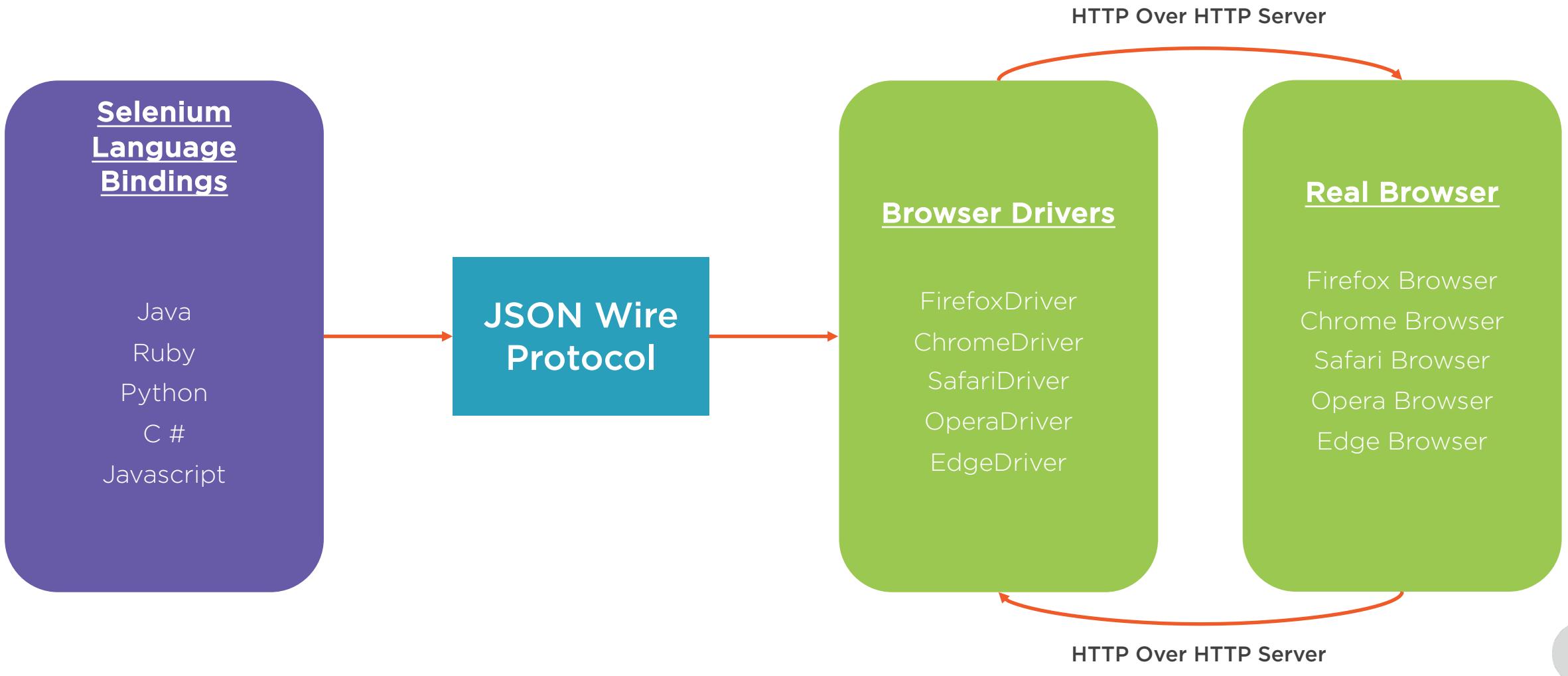
Selenium 1.0 + WebDriver = Selenium 2.0

Developed to better support dynamic web pages
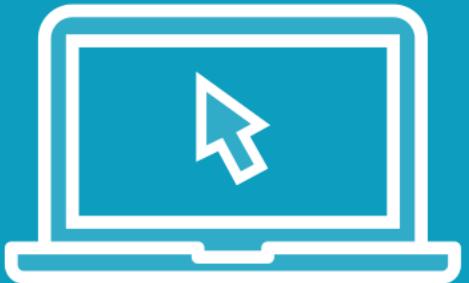
Browser specific classes

Selenium WebDriver Architecture

Demo
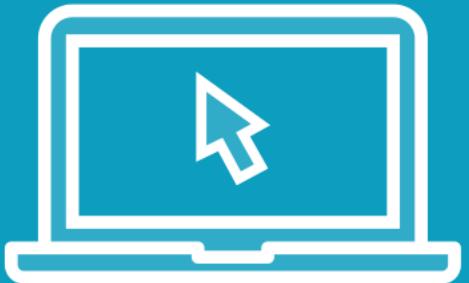
Demo : Installing required libraries and packages

1. Selenium
2. WebDriver
3. Beautiful Soup
4. Jupyter Notebook

Demo

Demo: Opening a webpage with WebDriver

Summary

Overview of web scraping in the context of static and dynamic content

Brief look at Selenium, components and use cases

Went deeper into WebDriver

Installing packages and opening a webpage with WebDriver