Documentation: Setting Up Environment for Web Scraping with Pyppeteer

This documentation will guide you through setting up the necessary environment for web scraping using Pyppeteer, a Python library for controlling a headless version of the Chrome browser.

**Step 1: Update Yum Package Manager**

**Commands:**

“sudo yum update –y”

This command updates the Yum package manager to ensure you have the latest versions of all packages.

**Step 2: Install Python3, Python3-pip, Node.js, and npm**

**Commands:**

“sudo yum install -y python3 python3-pip nodejs npm”

This command installs Python3, Python3-pip, Node.js, and npm which are required for running Pyppeteer and other dependencies.

**Step 3: Install Pyppeteer**

**Commands:**

“pip install pyppeteer”

Use pip, the Python package manager, to install Pyppeteer, a Python library for controlling headless Chrome using the DevTools protocol.

**Step 4: Download and Install Google Chrome**

**Commands:**

“sudo wget https://dl.google.com/linux/direct/google-chrome-stable\_current\_x86\_64.rpm”

“sudo yum localinstall google-chrome-stable\_current\_x86\_64.rpm –y”

Download the latest stable version of Google Chrome RPM package and install it using the yum package manager.

**Step 5: Install Required Dependencies**

**Commands:**

“sudo yum install libXcomposite libXdamage libXrandr libgbm libxkbcommon pango alsa-lib atk at-spi2-atk cups-libs libdrm”

Install additional dependencies required for Google Chrome and Pyppeteer to function properly.

Note: Ensure that the commands are executed in the correct order and without errors. Additionally, verify that your system meets the requirements for running headless Chrome and Pyppeteer.

With these steps completed, you should have a functional environment for web scraping using Pyppeteer.

**Step 6: Run the script**

**Command:**

python3 scrapper.py https://fb.com/bayyinahinst