### Overview

This Python script uses Streamlit to build a web app for querying websites. It utilizes Selenium for scraping, AmazonCaptcha for handling CAPTCHAs, and LangChain for text processing and querying.

### Installation

Install the necessary packages with:

See attach requirements.txt file

### Key Components

1. detect\_and\_solve\_captcha(driver): Detects and solves CAPTCHAs using Selenium.
2. Process\_urls(url, chrome\_driver\_path=None): Fetches and processes webpage content, handling CAPTCHAs if present.
3. process\_user\_query(question, vector\_db, chat\_history): Uses LangChain to answer user queries based on processed content.
4. main(): Configures the Streamlit app for URL input and query handling.

### Usage

**Run the App**:

streamlit run your\_script\_name.py

* **Configure URL**:
  1. Input the URL and (optionally) ChromeDriver path.
  2. Click "Submit URL" to process the webpage.

**Ask Questions**:

* 1. Enter your question in the provided text box to get a response.

### Issues and Findings

* **Chat History**: Encountered issues with ConversationBufferMemory not storing chat history when used with ConversationalRetrievalChain. Resolved by manually creating and managing a chat\_history list, appending each prompt and response.
* **CAPTCHA on Amazon**: Struggled with CAPTCHA solving on Amazon's website. Despite efforts, the CAPTCHA solving method using AmazonCaptcha was not successful. Manual CAPTCHA resolution might be necessary.

### Notes

* **Ensure ChromeDriver is compatible with your Chrome version.**
* **Handle CAPTCHA manually if automatic solutions fail.**