Bonus Task : Bypassing Amazon Captcha

Summary :

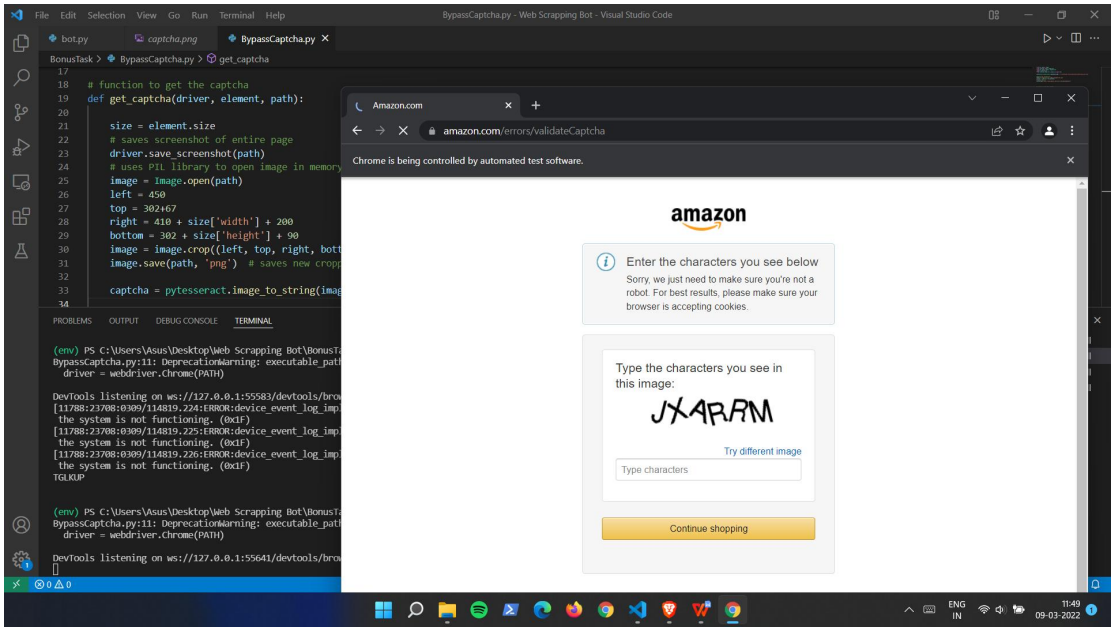
In this task I am going to take the screenshot of the amazon captcha webpage and crop the webage to get the png image of the captcha field . I am croping the image using pillow library and processing the characters in png file using pytesseract library. Then script send the extracted text to the captcha web page.

onusTask > BypassCaptcha.py > ...

```python
from PIL import Image
from selenium import webdriver
from PIL import Image, ImageFilter
import pytesseract
from selenium.webdriver.common.by import By

pytesseract.pytesseract.tesseract_cmd = r'C:\\Program Files\\Tesseract-OCR\\tes

#defing the chromedriver
PATH="C:\Program Files (x86)\chromedriver.exe"
driver = webdriver.Chrome(PATH)
#loading the url
driver.get("https://www.amazon.com/errors/validateCaptcha")




# function to get the captcha
def get_captcha(driver, element, path):

    size = element.size
    # saves screenshot of entire page
    driver.save_screenshot(path)
    # uses PIL library to open image in memory
    image = Image.open(path)
    left = 450
    top = 302+67
    right = 410 + size['width'] + 200
    bottom = 302 + size['height'] + 90
    image = image.crop((left, top, right, bottom))  # defines crop points
    image.save(path, 'png')  # saves new cropped image
```

```python
# function to get the captcha
def get_captcha(driver, element, path):

    size = element.size
    # saves screenshot of entire page
    driver.save_screenshot(path)
    # uses PIL library to open image in memory
    image = Image.open(path)
    left = 450
    top = 302+67
    right = 410 + size['width'] + 200
    bottom = 302 + size['height'] + 90
    image = image.crop((left, top, right, bottom))  # defines crop points
    image.save(path, 'png')  # saves new cropped image

    captcha = pytesseract.image_to_string(image)

    print(captcha)
    return captcha



img = driver.find_element(By.XPATH, "//div[@class='a-row a-text-center']/img")
captcha=get_captcha(driver, img, "captcha.png")
#sending the extracted text to the captcha webpage
fillcaptcha=driver.find_element(By.XPATH, "//input[@id='captchacharacters']")
fillcaptcha.send_keys(captcha)

driver.quit()
```

Sometime character recognition wrongly extacted the string due to complaxity of captcha


BypassCaptcha.py - Web Scrapping Bot - Visual Studio Code
bot.py
captcha.png
BypassCaptcha.py
BonusTask > BypassCaptcha.py > get_captcha
# function to get the captcha
def get_captcha(driver, element, path):
size = element.size
# saves screenshot of entire page
driver.save_screenshot(path)
# uses PIL library to open image in memory
image = Image.open(path)
left = 450
top = 302+67
right = 410 + size['width'] + 200
bottom = 302 + size['height'] + 90
image = image.crop((left, top, right, bott
image.save(path, 'png') # saves new cropp
captcha = pytesseract.image_to_string(imag
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
(env) PS C:\Users\Asus\Desktop\Web Scrapping Bot\BonusTa
BypassCaptcha.py:11: DeprecationWarning: executable_path
driver = webdriver.Chrome(PATH)
DevTools listening on ws://127.0.0.1:55583/devtools/brow
[11788:23708:0309/114819.224:ERROR:device_event_log_impl
the system is not functioning. (0x1F)
[11788:23708:0309/114819.225:ERROR:device_event_log_impl
the system is not functioning. (0x1F)
[11788:23708:0309/114819.226:ERROR:device_event_log_impl
the system is not functioning. (0x1F)
TGLKUP
(env) PS C:\Users\Asus\Desktop\Web Scrapping Bot\BonusTa
BypassCaptcha.py:11: DeprecationWarning: executable_path
driver = webdriver.Chrome(PATH)
DevTools listening on ws://127.0.0.1:55641/devtools/brow
Amazon.com
amazon.com/errors/validateCaptcha
Chrome is being controlled by automated test software.
amazon
Enter the characters you see below
Sorry, we just need to make sure you're not a robot. For best results, please make sure your browser is accepting cookies.
Type the characters you see in this image:
JXARRM
Try different image
Type characters
Continue shopping
ENG
IN
11:49
09-03-2022