**資料科學 期末專案**

**410723015 資工三 盧俊銘**

**1.System name :** 台鐵訂票爬蟲實作

**2.Motivation:**

Why do you want to develop the system?

連假要搶票，所以想用課堂所學Python之爬蟲及自動化技術，來完成自動訂票動作。

只需要填好資料即可自動訂票。

**3. Purpose:**

要完成訂票系統，我以 selenium做自動化後台操作，將使用者輸入的內容透過selenium填寫至webdriver，並以tkinter作為GUI開發套件，但是台鐵的網站上有 Google Recaptcha V2 我不是機器人驗證，所以我使用2captcha的API來做自動驗證我不是機器人之驗證。

本系統主要開發功能為自動化後台操作及驗證我不是機器人為兩個主要的功能，在下文惠提到細節內容。

**4. Relative work:**

市面上有許多台鐵訂票的APP在手機APP STORE上，這類產品都是透過使用者事先輸入內容，再去進行訂票動作，但是絕大多數的APP都要使用者去輸入我不是機器人的驗證碼，之後才能去做訂票的動作，但是我透過2captcha 即可自動驗證 Recaptcha的驗證。

**5. System description:**

**Developed packages (or modules) and tools:**

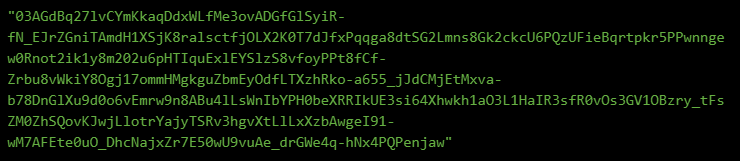
1. Selenium: 為瀏覽器自動化（Browser Automation）需求所設計的一套工具集合，讓程式可以直接驅動瀏覽器進行各種網站操作。許多 Web Test Framework，都是以 Selenium API 作為基礎，藉此操作網頁表單資料、點選按鈕或連結、取得網頁內容並進行檢驗。

2. Tkinter: 是一個python模組，是一個呼叫Tcl/Tk的介面，它是一個跨平臺的指令碼圖形介面介面。Tkinter不是唯一的python圖形程式設計介面，但是是其中比較流行的一個。

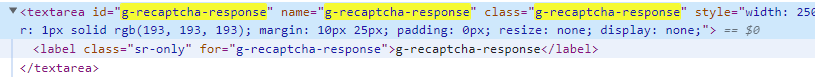
3. 2captcha: 是一家提供自動驗證碼識別服務的公司，主要提供一種人工圖像和CAPTCHA識別服務。 2Captcha的主要目的是由人類員工快速準確地解決您的驗證碼，但服務不僅限於解決驗證碼，還可以將人類可以識別的任何圖像轉換為文本。  
2Captcha是一個平台，通過鏈接需要實時識別大量驗證碼的用戶和需要掙錢的工人完成服務的交付。 2captcha可以識別任何當前已知的驗證碼。

🡨 2captcha運作流程圖

下圖為 2captcha 回傳之結果

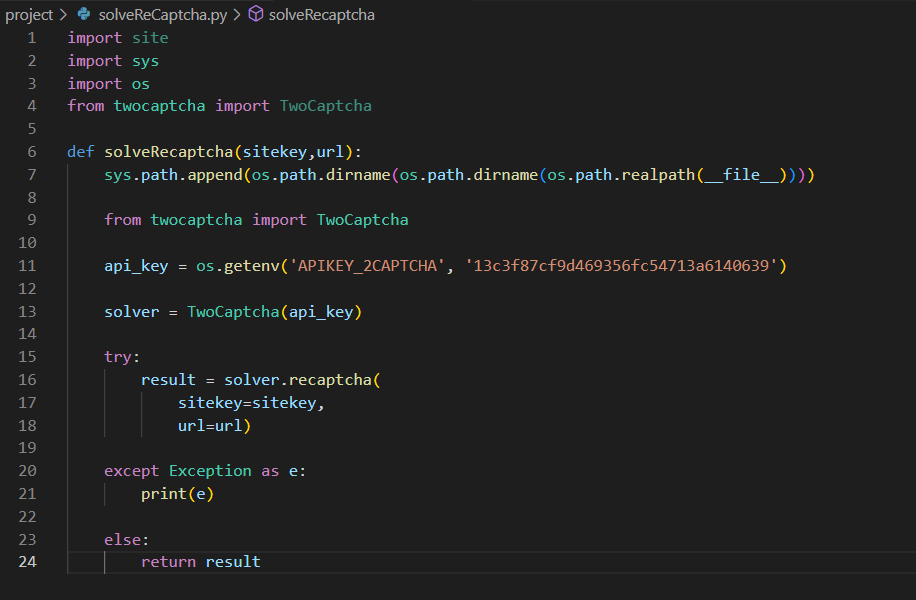


之後只要找到網站上textarea id = g-recaptcha-response 並填入收到的結果即可過驗證

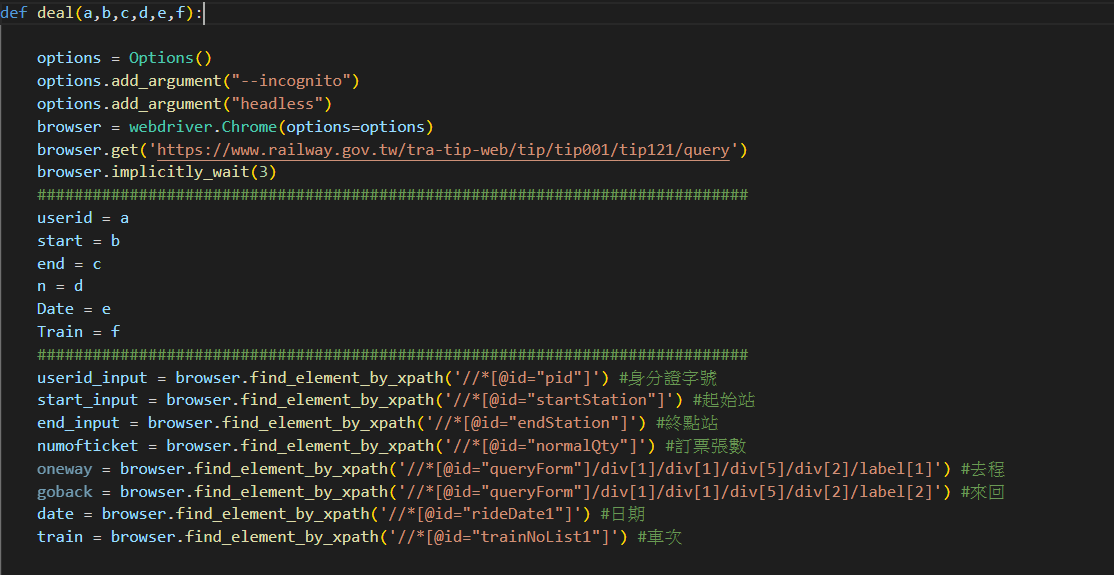


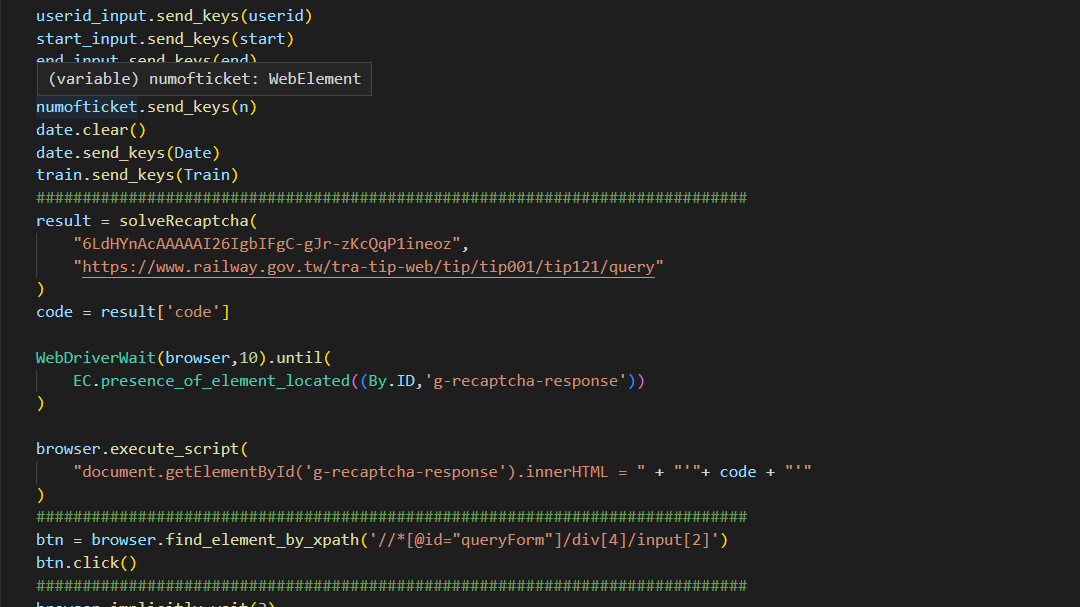
**System functions and main techniques:**

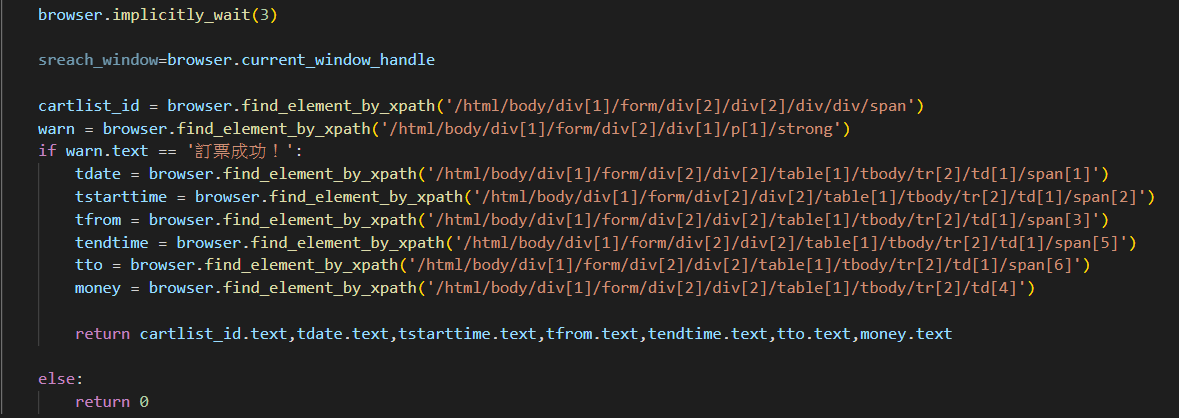
1. 驗證 我不是機器人 function



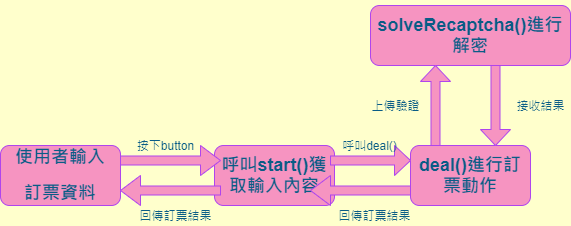
1. 訂票fumction







**the process of development**



How many hours do you spend on this system?

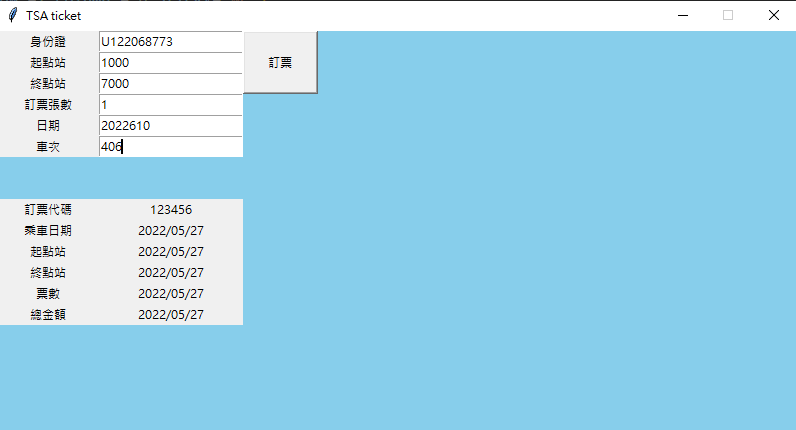
7~8 hrs.

What percentage have you developed by yourself for this system?

都是自己做的，但是有參考一些網路上的範例去完成這專案。

**6. Conclusion and reflection:**

下圖為訂票前

****

下圖為訂票成功

****

結論及反思:

專案可以如預期執行，但是目前缺點是由於2captcha是人工進行驗證程序的，他需要花費一定的時間去完成他(上傳及下載)，大概要1分鐘左右，這導致如果真的在做搶票時，並無法那麼效率的訂票，但是可以透過多開的方式，同時間進行多重搶票，這可大幅提升搶票機會，在未來有時間，我希望可以增加更多功能，如預約訂票、直接連結付款頁面等更便利的功能。