NLP LAB 新生作業題目敘述：

作業一

請撰寫一個程式來上網爬取教育部全球資訊網 即時資訊的資料

網址： <https://www.edu.tw/News.aspx?n=9E7AC85F1954DDA8>

程式語言建議使用python撰寫

網路爬蟲的部分可以使用jsoup或是beautiful soup等套件

我們要爬取的資料有 :

文章(即時新聞**[圖一]**): **時間、發布單位、標題、內文的聯絡人及電話[圖二]**

**1.手動輸入需爬取的文章的數量 [python語法 input()]**

**2.手動輸入需爬取的發布單位 [python語法 input()]**

**3.請設計防呆機制**

最後把撰寫程式的過程，編寫成一個教學投影片，如果自己在過程中遇到一些困難，也可以稍加解釋是如何解決的。

繳交檔案：

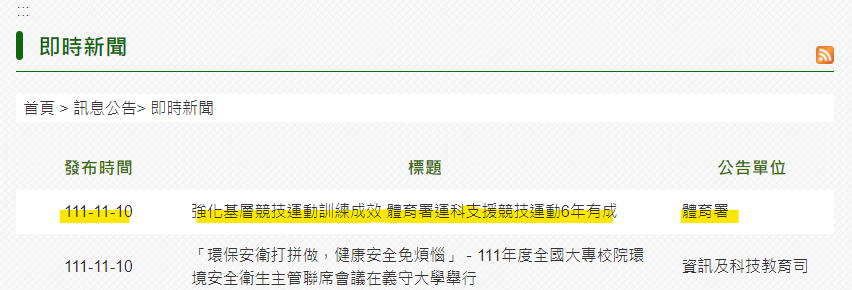
1. 三個json格式檔案。json欄位請依**[圖三]** (體育署、國教署、青年署各50筆資料)。

2. 一個教學投影片

3. 程式碼

以上請打包成壓縮檔，並以個人姓名為檔案名稱，附上信件回信即可。

若無法完成，也可以寫上自己無法解決的困難與未完成的程式。



圖一



圖二



圖三

作業二

Document1：

貓，通常指家貓（學名：Felis catus或Felis silvestris catus），為中型貓科動物。根據遺傳學及考古學分析，人類養貓的紀錄可追溯至10,000年前的新月沃土地區，古埃及人飼養貓的紀錄可追溯至1000年前，以防止老鼠吃掉榖物。現在，貓成為世界上最為廣泛的寵物之一，飼養率僅次於狗，但同時也威脅著很多原生鳥類、齧齒類的生存，是世界百大外來入侵種之一。長期飼養的貓平均壽命為12年以上，歷史上最長壽的貓則達38歲。

Document2：

犬（學名：Canis lupus familiaris），現代俗稱為狗，一種常見的犬科哺乳動物，與狼為同一種動物，生物學分類上是狼的一個亞種。狗是人類最早馴養的一個物種。有人飼養的稱為家犬，無人飼養的狗稱常被稱為「野狗」或「流浪狗」，可能對環境與其他生物造成影響，是世界百大外來入侵種之一。狗的壽命最多可超過二十年，平均則為十數年，與貓的平均壽命相近。若無發生意外，平均壽命以小型犬為長。

今天有一個問題：什麼動物是世界百大外來入侵種之一而且是狼的一個亞種?

1. 哪篇文章能回答這個問題 ?
2. 如果從電腦的角度，你會怎麼判斷哪篇文章能回答這個問題 ? (自由發揮，無正確答案)

作業三

A

1.如果今天你是一個科技業的老闆，在chatgpt還沒被推出前，你會運用自然語言的技術推出怎麼樣的產品或服務 ? 它的功能和特別之處是什麼 ?

2.在chatgpt推出後，你會怎麼整合chatgpt來為公司帶來更多的效益 ?

B

1.你覺得chatgpt在未來可能會有什麼樣的發展 ?