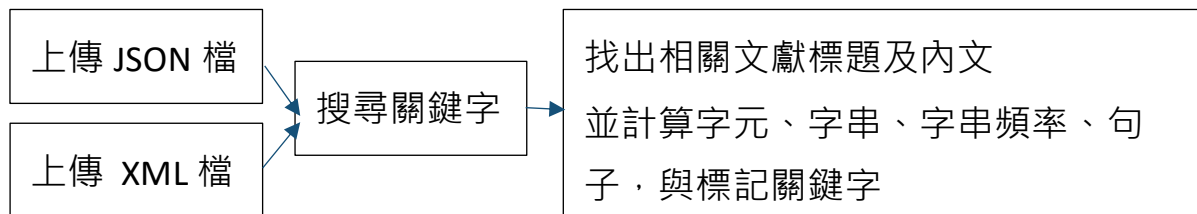


使用工具：前端(Html+Js)

後端(Python)

### 一、程式流程：



前端部分採用 JSON 和 XML 分開方式上傳，上傳後進入同一個關鍵字搜索，判斷檔名(JSON 或 XML)並對他們分別做資料處理，最後相關內容回覆在網頁上。

### 二、資料處理階段：

拆解：**XML**--使用 **ElementTree** 對來自 **pubmed** 資料的 ID、Article、Abstract 做拆解；**JSON**--使用 **JSON** 對 **Twitter** 資料的 ID、Text 做拆解。

文件分析：**Pubmed(XML 格式)**，撈下來的資料有些有缺摘要的標籤--只要有標題都做保留；**Twitter(JSON 格式)**，撈下來的資料有缺陷(JSON 間缺少逗號、有些是刪除的資料、文件內容語言問題...等) --把刪除的資料及非英文類別除，最後把逗號補齊。

### 三、相關文獻顯示：

字元、字串、字串頻率：使用 **split**、**dict**、**len** 算出

句子：對文件做適合他的判斷方式，**Pubmed(XML 格式)**—書寫拘謹，因此對其做句號、問號、驚嘆號等拆解；而 **Twitter(JSON 格式)**為聊天式，句子較不拘謹，因此用空行方式、**TAG 模式**做判斷。

標記關鍵字：使用 **replace** 更改關鍵字的 **CSS**

### 四、問題討論：

句子判斷仍不太準確，還需加強  
理想能做到不跳頁的顯示方式呈現