**IIR HW 1 方郁文**

**使用說明：**

上傳檔案

localhost:8000/search\_engine/upload\_file

搜尋文件

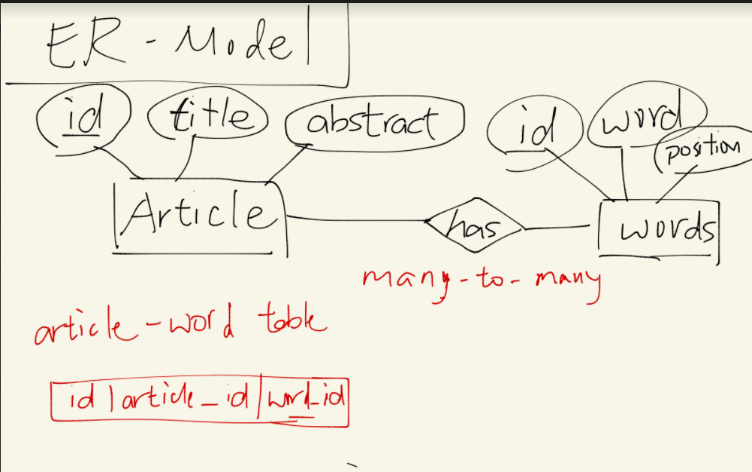
localhost:8000/search\_engine/search\_page

顯示目前所有文件

localhost:8000/search\_engine/show\_articles

**作法說明：**

建立database：



將上傳的文檔abstract存入資料庫中的Article資料表

將解析後的文字存入Word資料表，而文字和文檔的關係為many-to-many

解析文字：

1. Tokenization

以空白分割，將整個段落先分為許多tokens，透過python內建str.split()實作

1. Normalize

將大小寫統一，透過python內建str.lower() 實作

1. Stopwords

定義stopwords，並將其排除在外

1. Stemming

使用Porter Stemming Algorithm將文字做stemming後，存入Word資料表