**IR作業四 - PLSA**

資工碩一 M10715052 林仲鎧

心得：

這次的作業聽到要利用第一次作業的程式來做，剛聽到的時候我臉都綠了，第一次作業由於只求通過baseline，因此我就隨便亂刻了一個勉強及格的程式，但是這次是會有競賽成分存在的給分，所以我也是只能重刻一個程式來交這次的作業。

然而，事情並沒有想像的那麼順利，雖然程式很快地就重新寫好了，但是繳交上去的分數，竟然只有0.00001，我上IR課程到現在甚至沒看過這麼離譜的分數，而且不管我怎麼改，分數一直都是那麼低，所以真的是抓不到頭緒。

有一次，我問朋友這個作業是怎麼搞的，他只跟我說這個做業超簡單，後來我才發現，其實不能用老師第一章投影片中推薦的那三個方法，用最初始的TF-IDF的定義來做就好，所以我就把一些有的沒的加工算法都刪掉，沒想到刪掉之後分數直衝0.53，跨過baseline，讓我非常的驚訝。

這次作業由於query的量比較多，所以我還是用GPU來幫我進行運算，效益非常的好，從原本CPU運算需要600秒的時間，GPU只要20秒，足足有了30倍的加速，感覺2080Ti沒有白買！

最後在這裡附上我的參數，總共更新了6次，然後每次取top-13個，至於alpha跟beta的比率分別為0.15跟0.85。

