**IR Final Competition 心得**

組員:

M10715004 王德修 M10715007 周聿晟

M10715017 陳彥勳 M10715052 林仲鎧

論文參考: Khennak, Ilyes, and Habiba Drias. "An accelerated PSO for query expansion in web information retrieval: application to medical dataset." *Applied Intelligence* 47.3 (2017): 793-808.

心得:

一開始之所以會找這篇論文，主要是因為我們其中一位組員最近聽別人介紹啟發式演算法，也因此因緣際會知道了PSO(Particle Swarm Optimization)演算法，然而在找論文的時候，實在是找不太到與PSO演算法有直接相關的應用方法，好像是因為PSO應用在資訊檢索上的效果不太好，但是最後各方參考之後，還是有找到一篇有關於APSO(Accelerated PSO)的方法，並決定以此當作最後Competition的論文。

首先這篇論文的方法，是建立在APSO上的Query expansion，可以看到下方圖1，從基本的IR系統，一開始利用給定的query來做TF-IDF及BM25運算後得到相關document排序，到中間的Query expansion，利用基本的IR方式，一樣是BM25來得到document排序，在利用前面的document中的字，來作擴展搜尋，如此一來利用這些擴展的Query來進行檢索，便可以得到跟相關文章類似的文章，提高與使用者之Query有關文章的關聯度，但我們參考的這篇論文則不以為然，作者認為Query expansion所作的擴展詞，不應該是以擴展最多的詞來作為新的Query，而是要找接近原本Query所適合的擴展詞，如此一來找出來的相關document才會是跟使用者真正想尋找的資訊是有關連的。

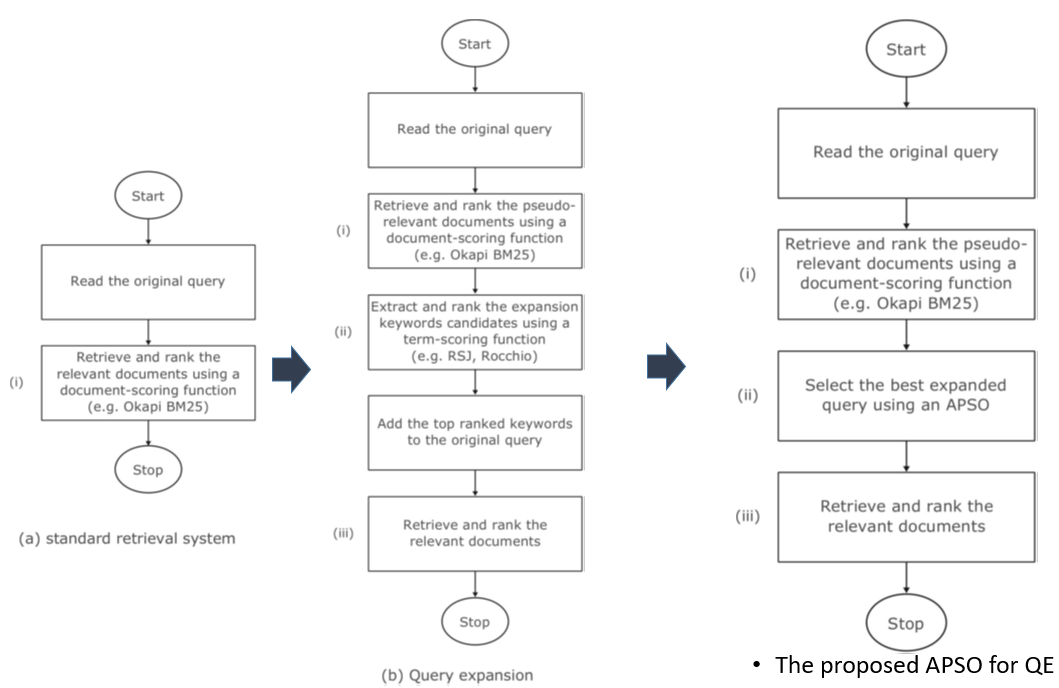


圖1

至於方法的部分，先來談述有關APSO跟PSO的差異，主要有三點，第一是APSO是加強了全域搜尋能力，是為了避免單純使用PSO會掉進局域最佳解的範圍內，第二是APSO擁有更快的收斂速度跟更好的穩定性，由於在資訊檢索領域中，對於不同種類的要搜尋的東西，用同一種方法得到的效果可能會更有優缺，APSO則是可以維持在一定的搜尋能力以上，最後，則是APSO也具有簡於實作和快速的執行效率的優點，如此一來可以在欲搜尋的資料庫非常龐大的時候，仍能保有在可接受的耗時中找出相關文章的能力。

接下來可以談論實作時的狀況，首先利用BM25的方法，來用原本的Query找到相關的document，並拿來排序，最後我們取K個排序較前面的document出來做Query expansion，然後決定擴展後的Query有幾個字，在讓取前K個document中的字來做隨機抽取，簡單來說，有點類似排列組合的感覺，只是在做出每一回的組合之後，會利用一個fitness function來替這個組合評分，並利用這個評分結果來得知，目前的組合與較好的組合之向量，如此一來便可以利用這個向量來進行移動，也就是選取更好的Query expansion組合，在這樣迭代個幾回之後，就能得到擴展完成的Query，並利用這個擴展完成的結果來做BM25即可以得到最終的結果。