Active Solid State Drives

主動式固態硬碟

組員:黃羿豪　　組別:A10　　指導老師:呂仁碩

摘要

固態硬碟(Solid State Drive)，簡稱SSD，是現在新興的資料存取裝置，具有低功耗、無噪音、抗震動、低熱量的特點，但是當資料庫增加的情況下，仍然有主機CPU處理時間增加的問題，因此主動式固態硬碟(Active Solid State Drives)的概念被提出，希望SSD亦可以負擔部分的CPU資料處理工作，能夠改善整體資料處理速度。

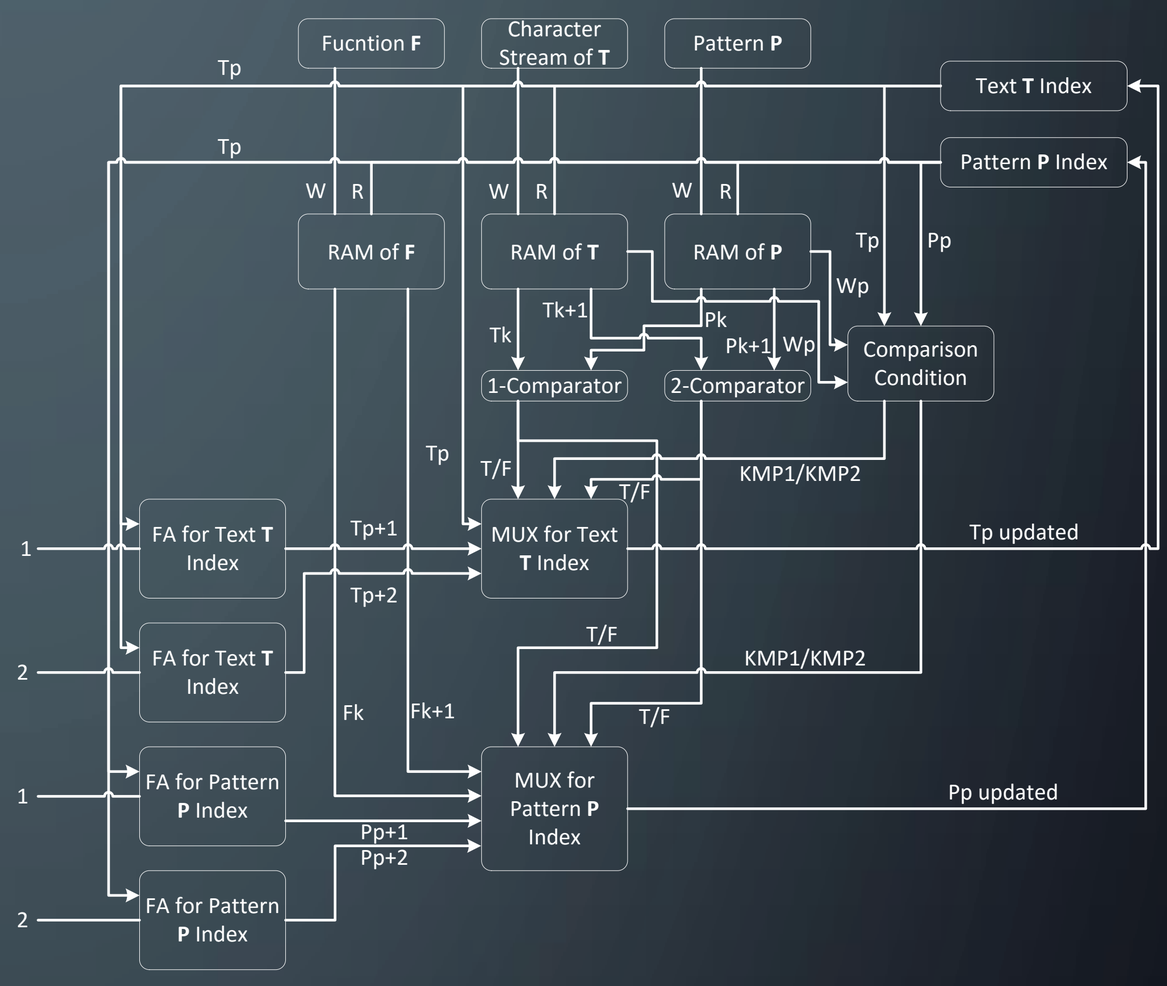
在資料庫分析的範疇中，字串比對(String Matching)有廣泛的應用，例如DNA序列比對，藉由字串模糊比對(Approximate String Matching)分析序列的相似程度，作為研究基因突變、遺傳的資料。

而專題以精確字串比對(Exactly String Matching)為導向，使用Knuth-Morris-Pratt Algorithm (KMP)字串搜尋演算法為基礎，在軟體上，最佳化用於分析字串結構的失敗函數；在硬體上，加入第二個比較器及緩衝器，允許硬體在每個週期都能寫入一個字元等待字串比對，希望將此硬體電路利用在Active SSD上，可以不須經過CPU的處理，透過Active SSD就可以有效的搜尋字串。

系統設計

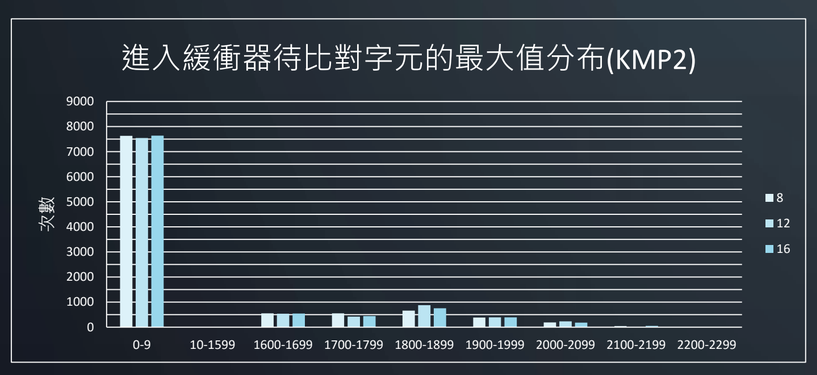
失敗函數F[N+1]定義：
Pattern P[N]與Text T[M]做字串比對
若 T_i 和 P_j 字元比對不相符時，$j = F_j$
若 T_i 和 P_{N-1} 字元比對相符時，$j = F[N]$

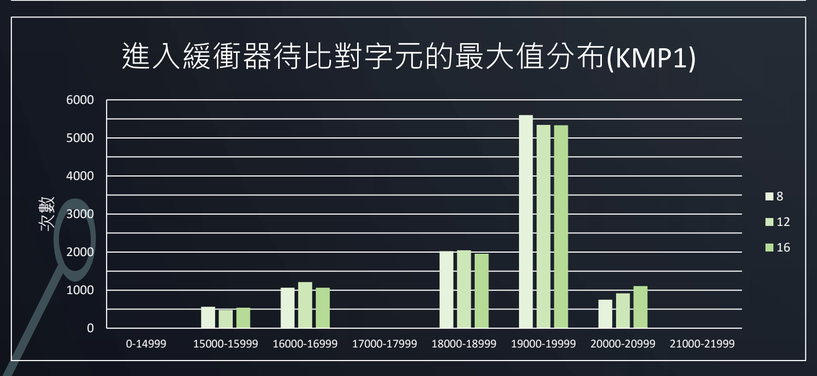
在硬體中，使用三個緩衝器分別存取 Pattern **P**、Failure Funftion **F**、Text **T**，將讀取 **T** 和 **P** 的字元在比較器 C1 、 C2 中做比對，根據讀取指標 Tp 、 Pp 和寫入指標 Wp ，判斷 C2 輸出是否視為隨意條件(Don't care condition)。多工器 MUX_T 、 MUX_P 以這些資料作為輸入選擇，輸出下一個讀取指標 Tp 、 Pp 。



實驗結果

模擬輸入DNA序列比對的資料，由ATCG字元組成的字串比對，以**P**長度為8、12、16，和一份長度為1ook的文件**T**，分別做10000組字串比對，可以發現所需要的緩衝器大小明顯下降許多，而且有許多字串只需要10個緩衝器以內，就可以允許硬體在每個週期都能寫入一個字元等待字串比對，當然比對完所有字元的週期也比原本KMP的演算法也下降許多。





結論

改善分析失敗函數以及利用兩個比較器做字元比對，的確改善了原本KMP演算法會因為比對錯誤，停止讀取新字元T_k的問題，但是在實驗中也發現，會有些字串**P**在新的演算法中，也是會需要超過1500個緩衝器才能不停止寫入，加以分析這些字串，發現還是有些特殊的結構組成，表示這個演算法仍然有待改進的空間。

參考資料

[1] Zachary K. Baker and Viktor K. Prasanna. Time and Area Efficient Pattern Matching on FPGAs