

CS305 作業系統概論 Prog. #2 Multithreading

2021.04.16

一、作業目的

熟悉如何使用Pthreads的API，撰寫multithreaded program，以及threads彼此之間如何共享資源。

二、作業內容

【言大師的生物奧秘探索】在生物奧秘的探索中，基因序列的搜尋，是最重要的基礎。在風之塔學院的言大師正走在時代尖端，探索生物基因序列的奧秘。在一個基因序列當中，常常會有一些特定的序列。如果用單執行緒的方式來尋找，程式會耗費許多時間，因此言大師需要一個程式，能夠利用多執行緒的方式來加速這個尋找的速度：

這個程式需要從命令列讀入一個資料檔案。檔案中包含三樣資訊：（1）一個長目標字元序列（最長為10kB，其開始位置為0），（2）整數n，用來表示每個搜尋序列要用幾個執行緒來平行搜尋，（3）以及數個需要尋找的字串序列。這些資料都由換行隔開。程式對每一個搜尋序列都要產生出 n 個執行緒來進行搜尋。為了加快速度，目標序列應該分成n份，來平行化搜尋。等所有的執行緒結束的時候，主執行緒要按由小至大的順序，列出每個搜尋序列的所有出現的起始位置，並印出整個程式的CPU時間。在產生執行緒的時候，要列印出（1）每個thread的id，以及（2）搜尋目標序列的起始位置與前8個字元。

例如：所輸入的資料檔如下

```
AACCGTTGGGCCCAAACACACCCACGGTT...TTTACCGGGAAACCCCTTTTTTTT
```

```
3
```

```
CCGTT
```

```
ACCGGGAAA
```

程式執行的過程可能是如此

```
$ prog2 prog2data.txt
```

```
[Tid=2048] search CCGTT at 0 AACCGTTG
```

```
[Tid=2049] search CCGTT at 100 CACCCACG
```

```
[Tid=2050] search CCGTT at 200 TTTACCGG
```

```
[Tid=2051] search ACCGGGAAA at 0 AACCGTTG
```

```
[Tid=2052] search ACCGGGAAA at 100 CACCCACG
```

```
[Tid=2053] search ACCGGGAAA at 200 TTTACCGG
```

```
[CCGTT] 2 50 251
```

```
[ACCGGGAAA] 68 232
```

```
CPU time: 105 ms
```

三、作業要點

1. 請注意，本作業使用的程式語言是C/C++，測試平台的作業系統： Ubuntu 20.04 64-bit。使用的編譯程式為gcc/g++ 編譯器： 9.3。其他平台或程式語言不在本次作業考慮範圍之內。如在測試平台上無法編譯與執行，都不予給分。
2. 請注意，本作業一定要用Pthread API來進行。任何不用Pthread API的程式，都不予給分。
3. 本作業的評分方式如下：
 - a. 基本功能：依照下面項目的完成程度來給分，如果只能完成部份，將部份給分。
 - i. 主執行緒能正確產生子執行緒。本項滿分20分。
 - ii. 主執行緒與子執行緒可以達成下面要求。本項滿分40分。
 1. 檔案只能讀一次，搜尋工作必須平均分配在所產生的執行緒。

2. 字元序列範圍是大寫英文字母{A,C,G,T}。
3. 目標序列最長為10K個字元，搜尋序列最長為32個字元。總共有2個搜尋序列。
4. n 為 2。（也就是，程式必須同時執行4個搜尋的子執行緒）
- iii. 子執行緒能用傳遞正確結果給主執行緒。本項滿分20分。
- iv. 子執行緒能自行用函式而不是透過主執行緒共用的變數拿到自己 `pthread_t` 的 `tid`。本項滿分10分。
- v. 主執行緒能正確印出CPU 時間，以毫秒為單位。本項滿分10分。
- b. 進階功能：完成以上基本功能且能得分達90分以上者，才可實作以下項目來得分。如果基本功能未達90分，進階功能的實作不予考慮。
 - i. 搜尋序列總共有2~5個搜尋序列。 n 在2~4之間。本項功能滿分10分。
 - ii. 在搜尋序列中，可以使用“?”表示一個萬用字元，例如：CC?TT，表示要搜尋以下四種序列：CCATT, CCCTT, CCGTT, CCTTT。完整完成本項功能可得15分。
 - iii. 在搜尋序列中，可以限定某個字元的範圍，字元的範圍以{}表示，例如：CC{A,C}TT，表示要搜尋CCATT及CCCTT。完整完成本項功能可得15分。
4. 本作業需繳交檔案：
 - a. 說明報告：檔案為docx或pdf格式。
 - i. 報告中必須說明程式的設計理念、程式如何編譯，以及**如何操作**。
 - ii. 報告中同時必須詳細說明你完成哪些部份。如有用到特殊程式庫，請務必說明。
 - iii. 請務必讓助教明白如何編譯及測試你的程式。助教如果無法編譯或測試，會寄信（**最多兩次**）通知你來說明，但每說明一次，**助教會少給你10分**。
 - b. 完整原始程式碼檔案（.c 或 .cpp）。**不可含執行檔。助教會重新編譯你們的程式。請注意**：不可用 .txt檔或是 .docx檔等非正常方式繳交程式碼，如有類似情形，**助教會扣10分**。
 - c. **不可以含有病毒，如果含有病毒等惡意程式，本作業0分。**
5. 所有相關檔案，例如報告檔、程式檔、參考資料等，請壓縮成一個壓縮檔（不可超過2MB）後上傳至portal。**請注意，不可抄襲。助教不會區分何者為原始版本，被判定抄襲者，一律0分。**

四、繳交方式：

1. 最終繳交時間：
 - a. 電子檔在 2021.05.17 以前，上傳至個人portal。如有多個檔案，將所有檔案壓縮成zip（rar 亦可）格式，然後上傳。
 - b. 上傳檔名格式：「學號_作業號碼.doc」或「學號_作業號碼.rar」。例如：912233_01.doc 或 912233_01.rar。
2. **如有違規事項者，依照課程規定處理。**
3. 如需請假，請上portal請假，並持相關證明文件，在請假結束後的第一次上課時完成請假手續，並在一週內完成補交。補交作業將以8折計算。
4. 老師不接受「門縫」方式繳交，助教也不接受任何作業。

五、如有未盡事宜，將在個人portal板面公告通知。

六、If you need **an English version** of this assignment or **any assistance in English**, please contact Prof. Yang.

七、參考資料

1. 參考課本圖 4.11。
2. PThread: <https://computing.llnl.gov/tutorials/pthreads/>

3. POSIX 線程 (pthread) 入門文章分享：<http://dragonspring.pixnet.net/blog/post/32963482-posix%E7%B7%9A%E7%A8%8B%28pthread%29%E5%85%A5%E9%96%80%E6%96%87%E7%AB%A0%E5%88%86%E4%BA%AB>
4. CPU時間的說明，<http://antoniohsu.blogspot.tw/2011/04/time.html>
 - a. 三種時間測量
 - 1) 實際時間(real time): 從command命令列開始執行到運行終止的消逝時間；
 - 2) 使用者CPU時間(user CPU time): 命令執行完成花費的用戶CPU時間，即命令在用戶態中執行時間總和；
 - 3) 系統CPU時間(system CPU time): 命令執行完成花費的系統CPU時間，即命令在核心態中執行時間總和。
 - b. CPU時間 = 使用者CPU時間 + 系統CPU時間
5. C/C++ 語言測量時間函數，評估程式執行效能方法整理：<https://blog.gtwang.org/programming/measure-the-execution-time-in-c-language/>