

108 學年度第一學期 演算法 第四次作業

1. Problem: 編輯距離 (Edit Distance)

[Goal]: 將一篇文章作一編輯後更正

Input: 一個字典(說明格式)及一篇文章

Output: i. BK-tree 字典 (Dic1, Dic2, Dic3)

ii. 根據 BK-tree 字典更正後的文章

iii. 分析

說明: i. 根據自行收集的英文單字建立字典(至少 200 英文單字)

如: 動物名稱(<https://tw.blog.voicetube.com/archives/44177>)、國家城市名稱(<https://tw.blog.voicetube.com/archives/39547>), 可自行設定主題。撰寫建構 BK-tree 作 search

ii. Edit distance: 實作三種 Edit Distance

a. 傳統 Edit Distance: Substitution 設為 1

b. Levenshtein Edit Distance: Substitution 設為 2

c. Damerau-Levenshtein Distance

(https://en.wikipedia.org/wiki/Damerau%E2%80%93Levenshtein_distance)

iii. 比較三種編輯距離的差異

iv. 比較三種編輯距離所建構的 BK-tree (要說明 BK-tree 存在檔案的結構)以及搜尋字串的比較次數(錯誤容忍值)

v. 輸出更正錯誤的字數(Word)與英文字(Character)的數量, 以及整篇文章字數(Words)和字元數(Characters)的錯誤率

2. 報告封面格式如下:

報告 title: 國立臺南大學資訊工程學系 108 學年度第一學期

課程 title: 演算法

作業: 第四次作業

報告名稱: 編輯距離 (Edit Distance)

班級: 資 XX

學號: S1XXXXXXXX

姓名: XXX

日期: 108 年 12 月 24 日

3. 繳交檔案分別為

- i. Edit-Distance_Report_學號.doc (完整報告)
- ii. Create-BKtree_學號.c + .exe (創建字典 programs, C, C++, ...)
- iii. Edit-Distance_Prog_學號.c (Edit-Distance 問題 programs, C, C++, ...)
- iv. Edit-Distance_Prog_學號.exe (程式執行檔)
- v. Input: Dic1.txt, test1.txt, test2.txt, ... (測試檔 test? 至少兩個)
- vi. Output: BKtree-Dic1.txt, BKtree-Dic2.txt, BKtree-Dic3.txt

4. 報告內容完整性 (格式如附加檔):

- i. 簡介及問題描述
- ii. 理論分析
- iii. 演算法則

- iv. 程式設計語言、工具、環境與電腦硬體等規格說明
 - v. 程式 (含 source code, input code, and output code)
 - vi. 執行結果與討論 (執行時間等問題討論)
- 完整報告與程式(source code, input code, output code, and executable code)經壓縮後(RAR or ZIP)，依規定時間內上傳至教學網站。
5. 作業繳交日期: **108/12/24 下午 23:59 前 (deadline)**
6. 機上驗收日期: **108/12/25 下午 1:00-5:00 (地點和驗收規範另行公布，驗收未到者視為未繳交該作業)**
7. 遲交以 0 分計算，除特殊情況外。