Reading Report:

The String-to-String Correction Problem

學號: r08922136

姓名: 黃新予

1. 問題定義  
   String-to-String Correction problem是要計算將一段文字利用3種編輯操作 (edit operation): 替換 (change), 刪除 (delete), 以及插入 (insert) , 轉變成另一段文字所需要的最少成本。  
     
   舉例，在三種編輯操作成本都一樣的情況下  
   將 “abc” 轉換成 “abd”，所需要的操作就是將c替換成d。  
   將 “abc” 轉換成 “ab” ，所需要的操作就是將c刪除。  
   將 “abc” 轉換成 “abcd”，所需要的操作就是在c後面插入d。
2. 解法敘述  
   解上述的問題可以用top-down的方式去解，例如要將string A 轉成 string B，其中A有i個字元，B有j個字元。  
     
   要使A[1…i] 轉換成 B[1…j] 有最小的cost，那最後一步也會經過3種編輯操作，也就是說 A[1…i] 變成 B[1…j] 是從以下取出最小值  
   a. A[1…i-1] 轉換成 B[1…j] 的最小成本，加上最後一步將A[i]刪除的成本;  
   b. A[1…i] 轉換成 B[1…j-1] 的最小成本，加上最後一步將B[j]插入的成本;  
   c. A[1…i-1] 轉換成B[1…j-1] 的最小成本，加上兩邊最後一個字元是否需編輯操作:  
    i. 如果A[i] 和 B[j] 相同，則不須經過任何編輯操作  
    ii. 如果A[i] 和 B[j] 不相同，則須將A[i] 替換成B[j]  
     
   所以將A[1…i] 轉換成B[1…j] 即是上述成本中，{a, b, c-i, c-ii} 中取最小值  
   (以上描述有參考此網站: https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/431735/)  
     
   接著我們可以仿照上述步驟，展開a, b, c中的最小成本，就可以一路回朔到初始狀態。  
     
   初始狀態為  
   a. A[0] 轉換成 B[0] 成本為0，因為兩邊都為空字串  
   b. A[0] 轉換成 B[1…j] 成本為 j \* 插入成本  
   c A[1…i] 轉換成 B[0] 成本為 i \* 刪除成本  
   利用以上資訊即可構成演算法。  
   舉例說明:  
   A: abc  
   B: bbe  
   假設3種編輯操作成本皆為1。  
   我們可以利用表格D呈現演算法過程，表格上方為A的字元，表格左側為B的字元。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| D | Φ | a | b | c |
| Φ |  |  |  |  |
| b |  |  |  |  |
| b |  |  |  |  |
| e |  |  |  |  |

初始狀態先填入

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Φ | a | b | c |
| Φ | 0 | 1 | 2 | 3 |
| b | 1 |  |  |  |
| b | 2 |  |  |  |
| e | 3 |  |  |  |

D[2, 2]: 讓A: a 轉換成 B: b的最小成本，即為  
a. A:空字串轉換成B: b的成本，再加上刪除a的成本 = D[2, 1] + 1 = 2  
b. A: a 轉換成B:空字串的成本，再加上插入b的成本 = D[1, 2] + 1 = 2  
c. A:空字串轉換成B:空字串的成本，再加上 a轉換成b的成本 = D[1, 1] + 1 = 1  
因此D[2, 2] = min {a, b, c} = 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Φ | a | b | c |
| Φ | 0 | 1 | 2 | 3 |
| b | 1 | 1 |  |  |
| b | 2 |  |  |  |
| e | 3 |  |  |  |

其他即可以此類推

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Φ | a | b | c |
| Φ | 0 | 1 | 2 | 3 |
| b | 1 | 1 | 1 | 2 |
| b | 2 | 2 | 1 | 2 |
| e | 3 | 3 | 2 | 2 |

因此A: abc 轉換成B:bbe的最小成本為 2，方式是將a, c 轉換成b, e。

1. 讀後心得  
   我覺得文中講到如果編輯操作的成本如果不是一致，而是根據情況而有所不同的話，此解法可以套用在拼音檢查工作上滿有趣的。  
   例如說如果我打字打 “qpple”, 一些文法工具就會挑出來問你是不是想打”apple”，以前我認為這個做法可能是跟資料庫去比對，但文中說到可以根據鍵盤上的每個英文字之間的距離去做校正，因為q和a的距離很近，因此由q替換成a的成本就會比較低，可以篩出更接近的詞彙。