107 學年度第一學期 演算法 第四次作業

1. Problem: **Plagiarism Detection based on Edit Distance**
   利用 Edit Distance 實作一抄襲偵測(Plagiarism Detection)系統。

   Functions:

   i. 利用 Dynamic Programming 實作 Minimum Edit Distance 演算法，包含 Minimum Edit Distances (一為 Insertion, Deletion, Substitution 皆各為一個 operation)以及 Levenshtein Distance (Insertion, Deletion,皆各為一個 operation，Substitution 為兩個 operations)兩種距離定義。
   a. Input: Strings S_1 and S_2 (利用 Keyboard 輸入)
   b. Output: Minimum Edit Distance (number) between S_1 and S_2

   ii. 利用(i)實作一抄襲偵測(Plagiarism Detection)系統，包含兩種 Minimum Edit Distances。
   a. Input: 原始檔案(S) and 比對檔案(T) (利用讀檔，宜考慮空白、Tab、Enter 等值)
   b. Output: 抄襲程度(各段以及全部檔案)

方法	Input	Output
(1)	原始檔案(S) … … 比對檔案(T) --- ---	兩篇文章抄襲程度： $d_{O,T} = 1 - \frac{\text{minED}(S,T)}{\max(\lvert S\rvert, \lvert T\rvert)}$ $d_{O,T}$=0, if $d_{O,T}$<0
(2)	原始檔案(S) 比對檔案(T) SP_1 SP_2 SP_3 SP_4 SP_n 共n段 TP_1 TP_2 TP_3 TP_m 共m段 d_{11} d_{12} d_{13} d_{1m}	兩段文字抄襲程度： $d_{i,j} = 1 - \frac{\text{minED}(SP_i, TP_j)}{\max(\lvert SP_i\rvert, \lvert TP_j\rvert)}$ $d_{i,j}$=0, if $d_{i,j}$<0 $k_i = \arg \text{MAX}_{1 \le j \le m} d_{i,j}$ **Note:** 要排除空白段，連續 Tab or 空白。 Outputs: a. 兩篇文章抄襲程度：$d_{S,T} = \frac{\sum_{i=1}^{n} d_{i,k_i}}{n}$ b. 段抄襲程度的變異數 c. 段抄襲程度的標準差

2. 繳交檔案分別為
   i. PD_Report_學號.doc (完整報告)
   ii. PD_Prog_學號.c (Plagiarism Detection 程式 c, c++等)
   iii. PD_Prog_學號.exe (Plagiarism Detection 執行檔)
   iv. TestFile1.txt & TestFile11.txt, TestFile12.txt, … (Test 1 set)
   v. TestFile2.txt& TestFile21.txt, TestFile22.txt, … (Test 2 set), …

3. 報告內容完整性：
   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, test files)
vi. 執行結果與討論 (功能討論、執行效率(時間)等討論)

- ❑ 完整報告與程式(source code, test files, and executable code)經壓縮後(RAR or ZIP)，依規定時間內上傳至教學網站。

4. **作業繳交日期： 107/12/26 23:59 前** (上傳至教學網站, It is hard deadline)

5. 遲交以 0 分計算，除特殊情況。

6. 下兩頁為一輸入的範例，同學可自行上網找英文文章範例，或者程式碼。繳交作業時，務必上傳至少兩組的測試範例組(每一組須包含一原始檔案&兩個擬被比對檔案當作範例)。

Input Files: (Example 1)

`TestFile1.txt` (原始檔案 S)

The computer learns from a huge database of four million videos from volunteers and paid-for market researchers in various emotional states and the algorithms are constantly updated and tested against real-world scenarios.
The next stage is to integrate voice analysis and other measures of physical wellbeing such as heart rate and hand gestures.

`TestFile11.txt` (比對檔案 T_1)

A computer model has been developed that can predict what word you are thinking of. The model may help to resolve questions about how the brain processes words and language, and might even lead to techniques for decoding people's thoughts.
Researchers led by Tom Mitchell of Carnegie Mellon University in Pittsburgh, Pennsylvania, 'trained' a computer model to recognize the patterns of brain activity associated with 60 images, each of which represented a different noun, such as 'celery' or 'aeroplane'.

`TestFile12.txt` (比對檔案 T_2) // in web site http://www.bbc.com/news/technology-34797189

The computer that knows what you're thinking

It is one of the advantages of being human that we can, when the time is appropriate, conceal our innermost feelings.
But that could be about to change as computers get better not just at recognising faces but also at understanding what the person behind it is actually thinking.
That is the project Daniel McDuff has been working on at the Massachusetts Institute of Technology's Media Lab, where he is developing emotion-reading computers.
It could eventually lead to machines that have emotional intelligence, or even everyday objects that are able to empathise with our moods - a mirror that knows how you feel about the way you look, a fridge that can offer you food that matches your state of mind or a car that recognises when you are stressed.
Joy and fear
The system Dr McDuff is developing works via a basic webcam that detects a range of different facial movements from frowning to smiling.
It translates that into seven of the most commonly recognised emotional states - sadness, amusement, surprise, fear, joy, disgust and contempt.

The computer learns from a huge database of four million videos from volunteers and paid-for market researchers in various emotional states and the algorithms are constantly updated and tested against real-world scenarios.
The next stage is to integrate voice analysis and other measures of physical wellbeing such as heart rate and hand gestures.
Cultural differences
Already the data has revealed that there are big differences in emotional responses between men and women and between different age groups and demographics.
"There are significant differences in different countries as to how people express themselves," Dr McDuff told the BBC.
"In collectivist cultures where your family group is more important, people are most expressive in these small groups, whereas in more individualistic cultures like Western Europe the emphasis is more about building relationships with strangers and people tend to be more positive around people who they are less familiar with."
He also found that when people mask their true feelings, the computer can recognise subtle differences.

Input Files: (Example 2)

TestFile2.txt （原始檔案 S）

```
#include <stdio.h>

void main()
{
    int a1, b2;
    a1=1;
    b2=2;
    printf("%d", a1+b2);
}
```

TestFile21.txt （比對檔案 T_1）

```
#include <stdio.h>

void main()
{
    int a, b;
    a=1;
    b=2;
    printf("%d", a+b);
}
```

TestFile22.txt （比對檔案 T_2）

```
#include <stdio.h>

void main()
{
    int a, b;
    printf("Hello World!");
}
```