**Problem 1：文章的用字統計**

本試題輸入檔為一篇中文文章，文章內容可能包括「中文」及若干「全型符號」。文章中可能

使用的「全型符號」包括「，」、「。」、「；」、「、」、「！」及「？」6 種。在中文中有若干

文字，如「臺」與「台」，在某些時候意義相同，但在本題中，只要是字碼不同，選手都應視

為「不同字」處理。

請由程式讀入給定的中文文章，並作以下的統計：

子題 **2.** （**8%**）

【使用路徑：**c:\Problem1\**子題**2\**】

【程式名稱：**p12**】

請選手在讀入的文章中，統計「出現次數最多」且「包含2 個中文字」的「詞」。在文章中，假設所有「兩兩相鄰的中文字」均可視為一個「詞」。本題所要統計的「詞」，只包括2 個中文字，其中不包含任何「全型符號」；我們也假設，一個「詞」不一定只出現在同一行中，也可能被切斷分隔在文章的前、後2 行。例如，有一篇文章內容為：

我們是學生。

今天要上學。

在此文章中，所有包括2 個字的「詞」有：「我們」、「們是」、「是學」、「學生」、「生。」「。今」、「今天」、「天要」、「要上」、「上學」及「學。」共11 組。雖然有些「詞」並不具明顯意義，如：「們是」、「是學」、「天要」及「要上」，但本題仍將其視為一種「詞」，並加入統計。但是「生。」、「。今」及「學。」等3 者，因為其中包括「全型符號」，則不算是「詞」，不列入統計。請選手統計並輸出出現次數最多、只包括 2 個中文字的「詞」。

輸入說明：

第 1 行表示文章的行數，第2 行開始為文章的內容。文章不超過20 行，每行不超過50 個中文字（含符號），而文章總字數不超過1000 字。

輸出說明：

若統計結果只有「1 個」最多出現次數者，則以1 行輸出「出現次數最多的『詞』」及其「出現次數」。但若有「多個」最多出現次數者，則以多行輸出。每一行輸出 1 個「詞」及其「出現次數」，輸出資料間以至少1 個空白隔開。

輸入檔案 **1**：【檔名：**in1.txt**】

8

為達成永續發展的目標，行政院規劃六年國家重點發展計畫，由教育部規劃的永續校園推廣計畫，為建立一個進步、安全、衛生、健康、人性化的學習環境空間為主，並於擴大內需刺激景氣的期程內，加速推行校園公共工程改造計畫，藉由突破傳統校園封閉的環境與制式管理原則下，整合社區共同意識、建立社區風貌、拓展生態旅遊等課題，改造校園環境成為具有社區特質的公共活動空間，結合校園綠色技術實施應用，轉化國內相關產業技術，進而增進綠色產業推廣效益，落實擴大內需進而促進產業升級與提振國內景氣之功效，從而發揮永續台灣、環境教育之積極意義與促成教育改革之目的。

輸入檔案 **2**：【檔名：**in2.txt**】

8

台灣地區資源短缺，興建大型水庫使用現階段推行不易，替代水源的之開發與再利用已成為未來資源開發的之重要途徑，推動節約用水，推廣再利用的中水系統，使全國平均之人均用水量逐步降至每日使用二百五十公升以下。落實國家水資源永續經營之理念，水利署推動雨水利用及生活雜排水回收再使用增加替代水源使用，並委託台灣雨水利用協會針對學界及各級學校相關人員，依各縣市所在之地理位置劃分北中南三區各辦理一場次雨水貯集使用及生活雜排水再利用講習演討會議，將邀請國內相關領域有豐碩成果與使用經驗之專家學者擔任主講員，對國內教育單位相關人員，以水資源永續利用、雨水貯集利用、生活雜排水再使用及節約用水等議題進行講習與研討。

輸出檔案：【檔名：**out.txt**】

校園 5

利用 7

使用 7

程式碼：

Imports System.IO

Public Class Form1

Dim fr As New FileInfo("in1.txt")

Dim frr As New FileInfo("in2.txt")

Dim fw As New FileInfo("out.txt")

Dim sw As StreamWriter = fw.CreateText

Dim ia, o2(2), x As Integer

Dim ib, o1(2) As String

Private Sub Form1\_Load(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles MyBase.Load

If fr.Exists = False Or frr.Exists = False Then

End

ElseIf fw.Exists = False Then

fw.Create()

End If

Dim sr As StreamReader = fr.OpenText

Dim srr As StreamReader = frr.OpenText

For i = 0 To 1

If i = 0 Then ia = sr.ReadLine

If i = 1 Then ia = srr.ReadLine

ReDim o1(2), o2(2) : x = 0 : ib = ""

For j = 1 To ia '將分段的文章全部變成同一行

If i = 0 Then ib &= sr.ReadLine

If i = 1 Then ib &= srr.ReadLine

Next

For j = 1 To Len(ib) - 1

If Asc(Microsoft.VisualBasic.Mid(ib, j, 1)) > -24000 And \_

Asc(Microsoft.VisualBasic.Mid(ib, j + 1, 1)) > -24000 Then '判斷兩個是否皆為中文字

For k = 0 To x '判斷是否有重複詞的

If Microsoft.VisualBasic.Mid(ib, j, 2) = o1(k) Then

o2(k) += 1 : Exit For

ElseIf k = x Then

o1(x) = Microsoft.VisualBasic.Mid(ib, j, 2)

x += 1 : ReDim Preserve o1(x), o2(x)

End If

Next

End If

Next

For k = 0 To x

For l = 0 To x '按數量大到小排序

If o2(k) > o2(l) Then

ib = o1(k) : o1(k) = o1(l) : o1(l) = ib

ib = o2(k) : o2(k) = o2(l) : o2(l) = ib

End If

Next

Next

For k = 0 To x

sw.WriteLine(o1(k) & "　" & o2(k) + 1)

If o2(k) <> o2(k + 1) Then Exit For

Next

sw.WriteLine()

Next

sw.Flush() : sw.Close() : End

End Sub

End Class