

2024年【科學探究競賽-這樣教我就懂】

普高組 成果報告表單

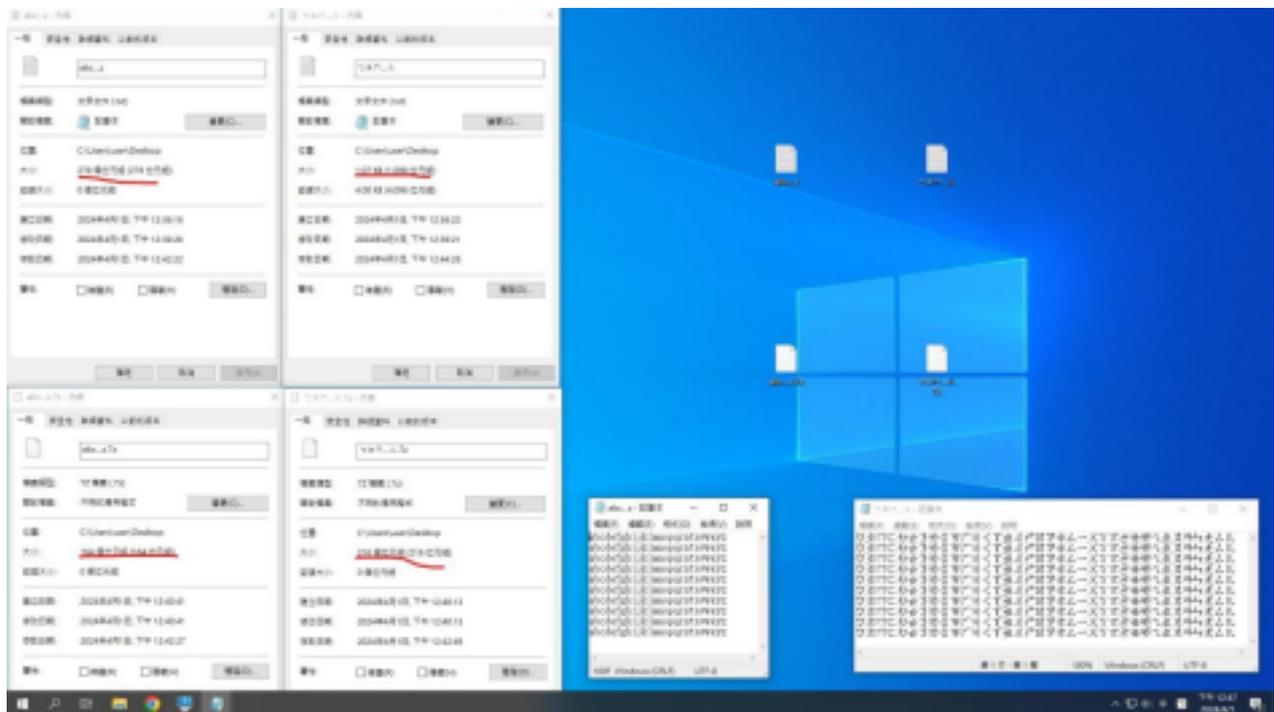
題目名稱：檔案壓縮 %&
一、摘要
ZIP 檔案是指通過壓縮軟體將多個檔案或資料夾壓縮成一個檔案的格式，以節省空間或方便傳輸。在壓縮的過程中，壓縮軟體會對選擇的檔案進行壓縮處理。這個處理的方式通常是通過使用壓縮演算法來減少檔案的大小，例如通過消除重複的資訊或使用更簡潔的編碼方式來表示資訊。這樣壓縮後的檔案就會佔用更少的空間，同時仍能夠通過解壓縮軟體還原成原始的檔案或資料夾。那麼其壓縮的程度有多少呢？
二、探究題目與動機
在使用雲端硬碟時，發現裡面的儲存空間快滿了，於是想到了可以用壓縮檔案的方式來減少儲存空間容量。當壓縮完後看減少了多少容量時，發現到有些檔案可以被壓縮很多，有些卻壓縮沒多少，於是就想知道為甚麼會出現這樣的狀況。
三、探究目的與假設
檔案內容：語言差異、檔案的資料冗餘性 檔案類型：.txt .jpg .mp3 是否可以被壓縮的更小？
四、探究方法與驗證步驟
觀察→假設→實驗→分析→結論 1.檔案內容： <ul style="list-style-type: none">• 語言差異：不同語言的文字在電腦上可能以不同的編碼方式存儲。某些語言的文字可能需要更多的位元組才能表示，而某些語言則可能需要較少的位元組。當壓縮包含多種語言

文字的檔案時，壓縮軟體會試圖找到一種通用的壓縮方式，以最小化檔案的大小。但有時

1

候，某些特定語言的文字可能導致壓縮效果不佳。

• 資料的冗餘性：檔案中可能包含重複的資訊或模式，這種情況稱為資料的冗餘性。壓縮軟體會嘗試找出這些重複的部分，並使用更簡潔的方式來表示它們，從而節省空間。然而，如果檔案中的重複資訊太少或者壓縮演算法無法有效地識別這些模式，那麼壓縮效果可能會受到影響。



上圖為中文和英文文字檔壓縮前與壓縮後差異，可知中文可以被壓縮部分的較多

2. 檔案類型：

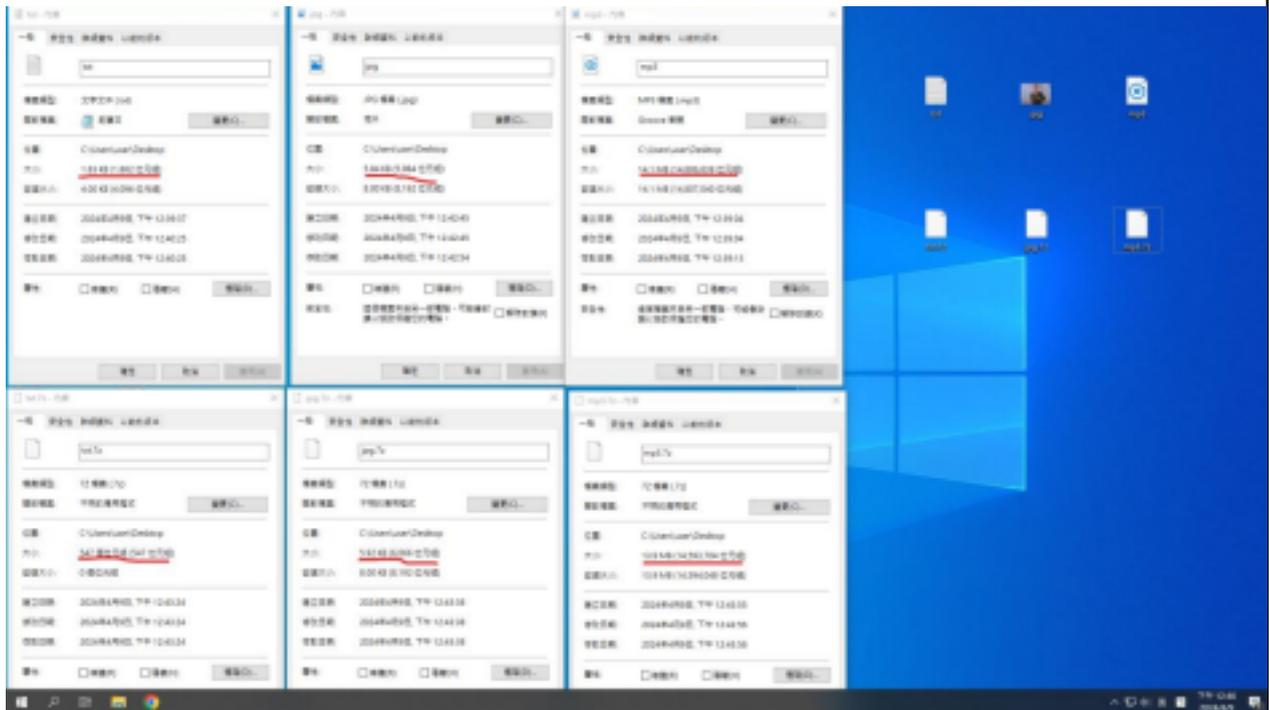
• 文字檔案 (.txt)：純文字檔案通常包含的是人類可讀取的文字內容，通常會有較高的壓縮效果。這是因為純文字檔案通常包含大量重複的字元序列，例如空格、換行符號等，壓縮軟體能夠較容易地識別和壓縮這些重複的內容。

• 圖像檔案 (.jpg)：圖像檔案通常具有較高的壓縮效果，因為它們包含的是圖像資訊，而非純文字。JPEG格式的圖像檔案使用有損壓縮演算法，這意味著一些細節和資訊可能會被刪

除或改變以達到壓縮目的。因此，JPEG格式的圖像通常能夠以較小的檔案大小來保存，但可能會導致一些細微的圖像品質損失。

2

• 音訊檔案(.mp3)：音訊檔案通常具有較高的壓縮效果，因為它們包含的是音訊資訊。MP3 格式的音訊檔案使用有損壓縮演算法，這意味著一些聲音細節和頻率可能會被刪除或改變以達到壓縮目的。因此，MP3 格式的音訊檔案通常能夠以較小的檔案大小來保存，但可能會導致一些聲音品質損失。



由上圖可知，.txt可被壓縮得很小、.mp3壓縮後比原本小了一點、.jpg卻變大了一點

五、結論與生活應用

在傳輸儲存大檔案，我們經常需要壓縮檔案以節省儲存空間或加快檔案傳輸速度。然而，經過發現，原來語言會影響檔案的大小。特別是，發現非英文的文字檔案能被壓縮得更多。這可能是因為非英文語言，如中文或阿拉伯文，使用的字符集比英文大，因此有更多的冗餘資訊可以被壓縮。此外也發現，不同類型的檔案壓縮後的大小差異很大。例如，文字檔案通常可以被壓縮得很多，這可能是因為文字檔案包含大量的重複字符，這些重複的資訊可以被有效地壓縮。然而，音檔的壓縮程度較小，這可能是因為音檔已經進行了一定程度的壓縮，因此無法進一步壓縮。令人驚訝的是，壓縮後的圖檔大小竟然會變大。這可能是因為圖檔通常已經進行了高效的壓縮，再次壓縮可能會增加一些元資訊，反而使檔案變大。這些發現提醒我們，檔案壓縮並非一種「一體適用」的解決方案，我們需要根據檔案

的類型和內容來選擇最適合的壓縮策略。

參考資料

<https://experience.dropbox.com/zh-tw/resources/how-to-zip-files>
<https://zh.wikipedia.org/zh-tw/ZIP%E6%A0%BC%E5%BC%8F>
<https://experience.dropbox.com/zh-tw/resources/what-is-a-zip-file>

註：

1. 報告總頁數以6頁為上限。
2. 除摘要外，其餘各項皆可以用文字、手繪圖形或心智圖呈現。
3. 未使用本競賽官網提供「成果報告表單」格式投稿，將不予審查。
4. 建議格式如下：
 - 中文字型：微軟正黑體；英文、阿拉伯數字字型：Times New Roman ● 字體：12pt為原則，若有需要，圖、表及附錄內的文字、數字得略小於12pt，不得低於10pt

- 字體行距, 以固定行高20點為原則
- 表標題的排列方式為向表上方置中、對齊該表。圖標題的排列方式為向圖下方置中、對齊該圖