**I. Kết quả thực nghiệm**

| Công cụ nén | Định dạng | Dung lượng sau nén | Tỷ lệ tiết kiệm |
| --- | --- | --- | --- |
| Không nén (Gốc) | - | 520 MB | 0% |
| Windows (ZIP) | .ZIP | 505 MB | 2.9% |
| WinRAR | .RAR | 495 MB | 4.8% |
| 7-Zip | .7Z | 485 MB | 6.7% |

**II. Phân tích và Nhận xét**

**1. Hiệu quả tiết kiệm dung lượng của từng công cụ**

* 7-Zip (.7Z) cho hiệu quả nén cao nhất (đạt tỷ lệ tiết kiệm 6.7%). Điều này là do 7-Zip sử dụng thuật toán LZMA/LZMA2 mạnh mẽ, được tối ưu hóa để tìm và loại bỏ các dữ liệu trùng lặp một cách hiệu quả, ngay cả với chi phí thời gian nén lâu hơn.
* WinRAR (.RAR) cho hiệu quả nén trung bình cao (4.8%). Thuật toán của RAR cung cấp sự cân bằng tốt giữa tỷ lệ nén và tốc độ xử lý.
* Công cụ tích hợp của Windows (.ZIP) cho hiệu quả nén thấp nhất (2.9%). Định dạng ZIP sử dụng thuật toán đơn giản, ưu tiên tốc độ nén nhanh chóng và tính tương thích rộng rãi hơn là tỷ lệ nén tối đa.

**2. Nhận xét về hiệu quả nén theo loại tập tin**

Quan sát quá trình nén, chúng tôi nhận thấy:

* Tài liệu văn bản (.docx, .txt) và File mã nguồn: Đây là loại file nén hiệu quả nhất. Cả ba công cụ đều giảm được đáng kể dung lượng (có thể lên đến 60-70%) vì chúng chứa nhiều văn bản thuần túy với các từ, ký tự lặp lại mà thuật toán nén không mất dữ liệu xử lý rất tốt.
* Video (.mp4) và Âm thanh (.mp3): Hiệu quả nén rất thấp, gần như không đáng kể. Lý do là các file này đã được nén mất dữ liệu (lossy) từ trước. Thuật toán nén không mất dữ liệu (ZIP, RAR, 7z) khó có thể tìm thấy sự dư thừa để nén thêm.
* Hình ảnh: Hiệu quả nén phụ thuộc vào định dạng ảnh:
  + Ảnh .JPG (đã nén lossy): Giống như video, nén lại gần như không giảm được dung lượng.
  + Ảnh .PNG (nén lossless): Có thể nén được một chút, nhưng vì bản thân PNG đã là định dạng nén, nên hiệu quả cũng không cao.

=> Kết luận: Đúng như dự đoán, tài liệu văn bản là loại file phản hồi tốt nhất với việc nén bằng các công cụ này. Video và âm thanh đã nén trước đó gần như "miễn dịch" với việc nén thêm.

**III. Giải thích lý do một số tập tin không thể nén nhiều**

Có hai nguyên nhân chính:

1. Tập tin đã được nén ở mức tối ưu (Pre-compressed):
   * Ví dụ: File .mp4, .mp3, .jpg.
   * Giải thích: Các định dạng này đã sử dụng thuật toán nén mất dữ liệu phức tạp để loại bỏ các thông tin mà mắt/ tai con người khó nhận biết. Dữ liệu còn lại sau khi nén gần như là ngẫu nhiên và không còn sự dư thừa để các công cụ như ZIP, RAR, 7z có thể nén được nữa. Việc nén lại đôi khi còn làm tăng nhẹ kích thước do phải thêm phần header của file nén.
2. Bản chất của dữ liệu không có tính lặp lại (Entropy cao):
   * Ví dụ: Các file đã được mã hóa, file thực thi (.exe).
   * Giải thích: Các thuật toán nén không mất dữ liệu hoạt động dựa trên nguyên tắc tìm kiếm và thay thế các chuỗi dữ liệu giống nhau. Nếu một file chứa dữ liệu ngẫu nhiên hoặc đã được mã hóa, nó sẽ không có các chuỗi lặp lại này, khiến cho việc nén trở nên vô hiệu.