Bài 3:

* Phân tích hiệu quả tiết kiệm dung lượng của mỗi công cụ. Đưa ra nhận xét về định dạng nén nào hoạt động tốt nhất với từng loại tập tin (ví dụ: video có thể nén mạnh hơn hình ảnh hoặc tài liệu văn bản).
* Giải thích lý do tại sao một số tập tin không thể nén nhiều hoặc không có sự khác biệt lớn về kích thước khi nén.

Trả lời:

* Phân tích hiệu quả tiết kiệm dung lượng của mỗi công cụ:

1. File văn bản, code, cơ sở dữ liệu (.txt, .csv, .docx, .xlsx):

* Hiệu quả nén rất cao
* Lý do: Dữ liệu dạng text có nhiều ký tự lặp lại, thuật toán nén dễ tối ưu
* Định dạng hiệu quả: .7z hoặc .rar

1. Hình ảnh (.jpg, .png, .gif):

* Hiệu quả nén thấp
* Lý do: Các định dạng này vốn đã dùng thuật toán nén (JPEG, PNG có cơ chế nén riêng)
* Định dạng nén không tạo khác biệt nhiều, ZIP/RAR/7z đều tương tự

1. Âm thanh và video (.mp3, .mp4, .avi, .mkv):

* Hiệu quả nén rất thấp
* Lý do: MP3 và MP4 đã được nén mạnh theo chuẩn lossy (mất dữ liệu). Công cụ nén gần như không thể giảm thêm
* Giải pháp: Muốn dung lượng nhỏ hơn thì phải chuyển đổi định dạng (re-encode) chứ không chỉ nén

1. Ảnh hoặc video chưa nén (.bmp, .wav, .raw):

* Hiệu quả nén cao
* Lý do: Đây là dạng dữ liệu thô, nhiều khoảng trống thông tin nên dễ nén
* Định dạng hiệu quả: .7z hoặc .rar

1. Phần mềm và file cài đặt (.exe, .dll, .iso):

* Hiệu quả nén trung bình
* Lý do: Một số file đã có cơ chế nén sẵn trong quá trình đóng gói
* WinRAR/7-Zip thường nén tốt hơn ZIP tích hợp của Windows
* Một số tập tin không thể nén nhiều hoặc không có sự khác biệt lớn về kích thước khi nén bởi vì :

1. Đã được nén sẵn: Các định dạng đa phương tiện như vốn đã tối ưu bằng thuật toán nén, nên nén thêm cũng không giảm đáng kể
2. Dữ liệu ngẫu nhiên hoặc đã được mã hóa: File đã được mã hóa hoặc chứa dữ liệu ngẫu nhiên thì không thể nén thêm
3. Độ lặp thấp trong dữ liệu: Thuật toán nén hoạt động hiệu quả khi dữ liệu có nhiều mẫu lặp lại. Với dữ liệu phân tán, đa dạng, nén sẽ không cải thiện nhiều