**BÁO CÁO ĐỒ ÁN**

BỘ MÔN: CS232 – Tính toán Đa Phương tiện

ĐỀ TÀI:

**THUẬT TOÁN NÉN LZW**

*Lempel – Ziv – Welch Compression Algorithm*

Sinh viên thực hiện: Hoàng Ngọc Bá Thi – MSSV: 19522255

Giảng viên hướng dẫn: Nguyễn Vinh Tiệp.

**MỤC LỤC**

1. **THUẬT TOÁN LZW**
2. **Giới thiệu về LZW**
3. **Thuật toán nén LZW**
4. **Thuật toán giải nén LZW**
5. **CÀI ĐẶT VÀ THỰC NGHIỆM THUẬT TOÁN LZW**
6. **Cài đặt thuật toán LZW**
7. Công cụ nén và giải nén văn bản
8. Công cụ nén và giải nén hình ảnh
9. **Thực nghiệm trên văn bản**
10. **Thực nghiệm trên hình ảnh**
11. **KẾT LUẬN**
12. **TÀI LIỆU THAM KHẢO**
13. **TÀI LIỆU ĐÍNH KÈM**

**PHẦN NỘI DUNG**

1. **THUẬT TOÁN LZW**
2. **Giới thiệu về LZW**

- Tên đầy đủ: Lempel-Ziv-Welch.

- Năm 1977, Abraham Lempel và Jacob Ziv lần đầu tiên đưa ra khái niệm “nén từ điển”. Sau đó phát triển thành một họ giải thuật nén từ điển LZ (LZ77, LZ78).

- Năm 1984, Terry Welch cải tiến LZ thành một giải thuật mới hiệu quả hơn, đặt tên là LZW.

- Phương pháp LZW dựa trên việc xây dựng từ điển cho các “chuỗi ký tự” đã từng xuất hiện trong văn bản, những “chuỗi ký tự” xuất hiện sau đó sẽ được thay thế bằng mã của nó trong bảng từ điển.

- Giải thuật LZW được sử dụng cho tất cả các loại file nhị phân. Nó thường được dùng để nén các loại văn bản, ảnh đen trắng, ảnh màu … và là chuẩn nén cho các dạng ảnh GIF, TIFF…

1. **Thuật toán nén LZW**

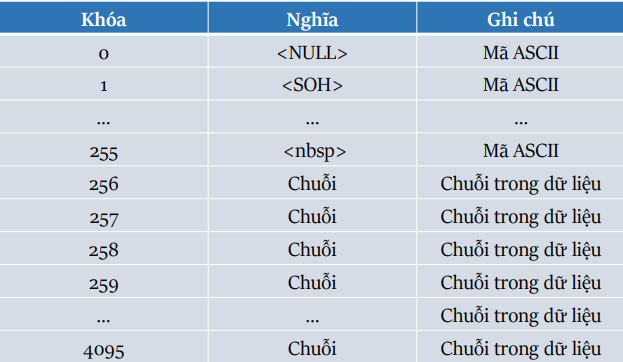
- Hoạt động theo nguyên tắc là tạo ra một từ điển động.

- Từ điển là tập hợp những cặp Khóa và Nghĩa của nó. Trong đó Khóa được sắp xếp theo thứ tự nhất định, Nghĩa là một chuỗi con trong dữ liệu.

- Từ điển được xây dựng đồng thời với quá trình đọc dữ liệu. Sự có mặt của một chuỗi con trong từ điển khẳng định rằng chuỗi đó đã từng xuất hiện trong phần dữ liệu đã đọc. Thuật toán liên tục tra cứu và cập nhật từ điển sau mỗi lần đọc một ký tự ở dữ liệu đầu vào.

- Người ta thường dùng từ điển với kích thước 4096 (212) phần tử.

* Cấu trúc từ điển có dạng như sau:

****

* Mã giả:

Cài đặt từ điển với các chuỗi ký tự đơn (Bảng mã ASCII 0-255)

P = Ký tự đầu tiên của chuỗi INPUT

**WHILE** chưa đọc hết chuỗi INPUT

C =ký tự tiếp theo trong chuỗi INPUT

**IF** chuỗi P + C có trong từ điển

P = P + C

**ELSE**

Xuất mã khóa cho P

Thêm P + C vào từ điển

P = C

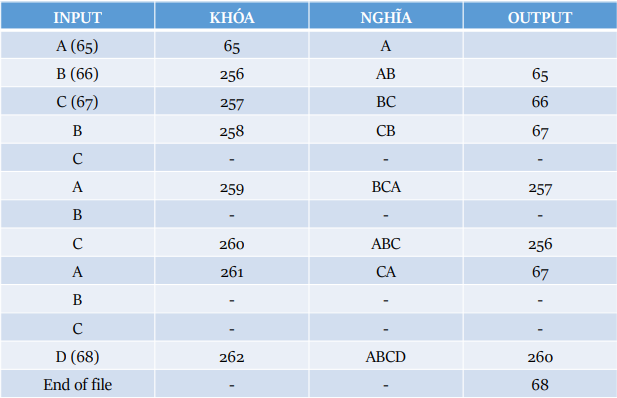
**END** **WHILE**

Xuất mã khóa của P

**Ví dụ:** INPUT: ABCBCABCABCD => 12\* 8bits = 96 bits.

OUTPUT CODE: 65 66 67 257 256 67 260 68 => 8 \* 9 bits = 72 bits.

Tỉ số nén: 72 / 96 = 0,75.

****

1. **Thuật toán giải nén LZW**

* Mã giả:

Cài đặt từ điển với các chuỗi ký tự đơn

OLD = Mã khóa đầu tiên trong dãy INPUT

Xuất ra bản dịch của OLD

**WHILE** chưa đọc hết dãy INPUT

NEW = Mã khóa tiếp theo trong dãy INPUT

**IF** NEW không có trong từ điển

S = bản dịch của OLD

S = S + C

**ELSE**

S = bản dịch của NEW

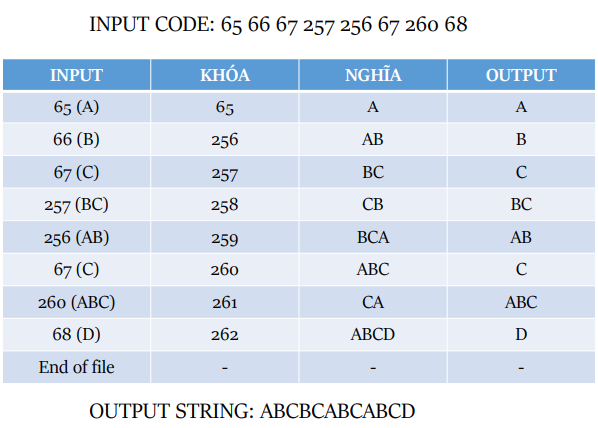
Xuất S

C = ký tự đầu tiên của chuỗi S

Thêm OLD + C vào từ điển

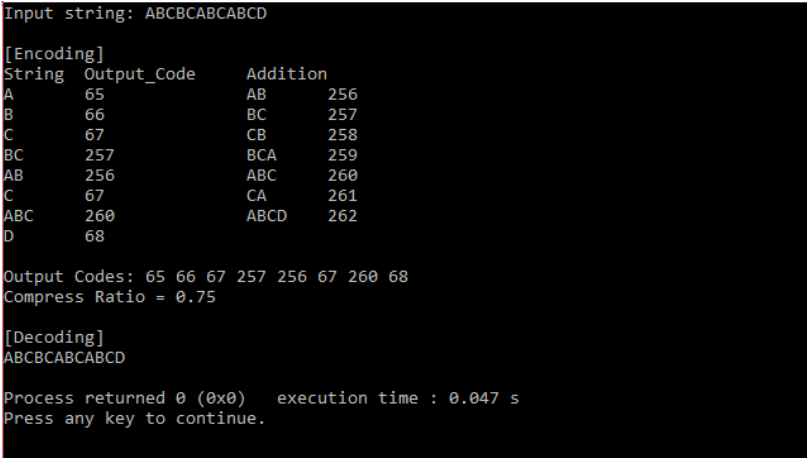
OLD = NEW

**END WHILE**

****

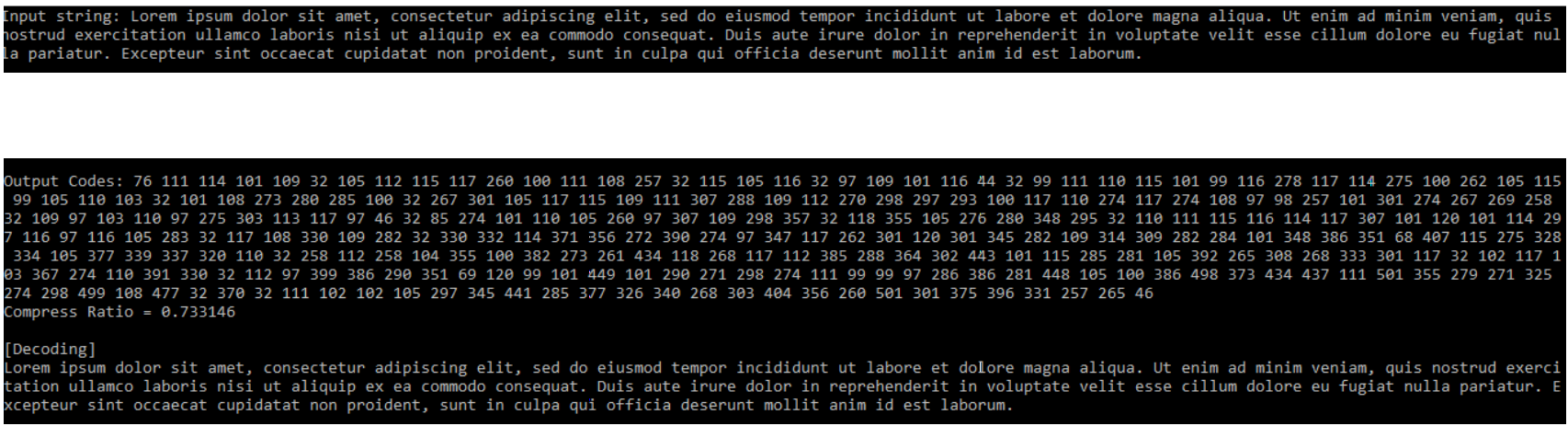
1. **CÀI ĐẶT VÀ THỰC NGHIỆM THUẬT TOÁN LZW**
2. **Thực nghiệm trên văn bản**

***Thực nghiệm 1*:** input nhỏ, có sự lặp lại các ký tự.



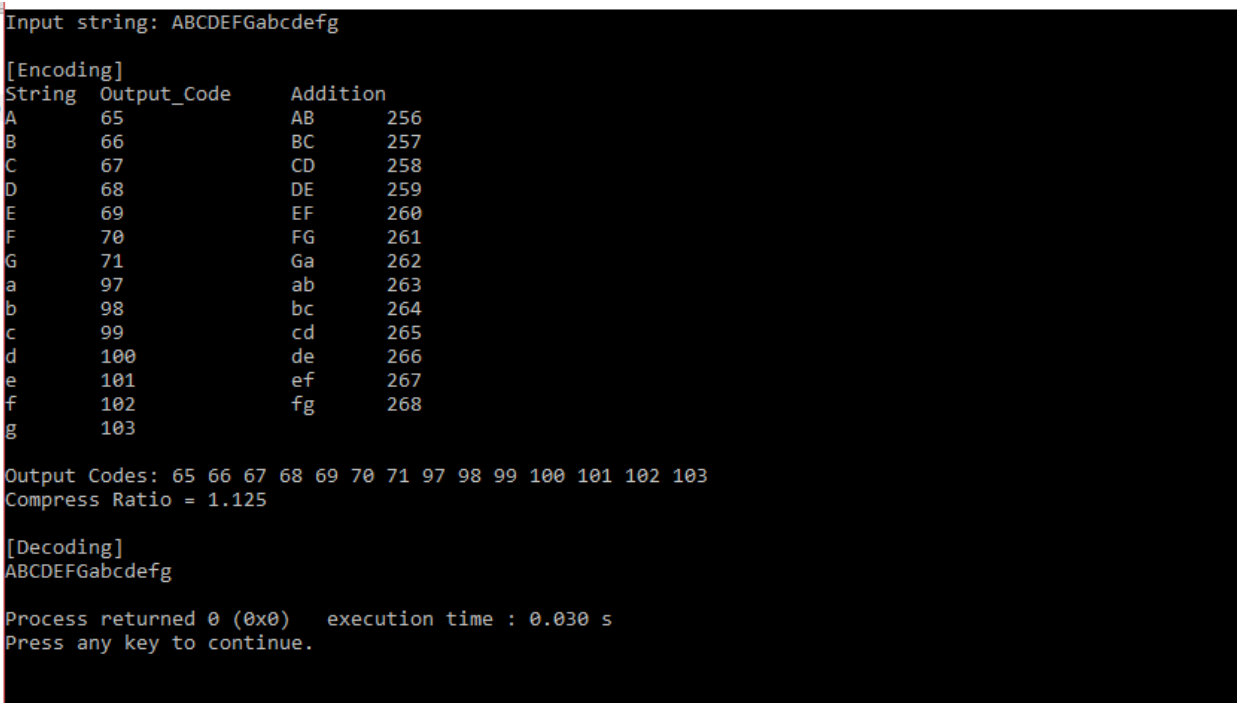
**Tỉ số nén: 0.75**

***Thực nghiệm 2***: input lớn, có sự lặp lại các ký tự.

****

**Tỉ số nén: 0.733146**

***Thực nghiệm 3:*** input không có sự lặp lại các ký tự.

****

**Tỉ số nén: 1.125**

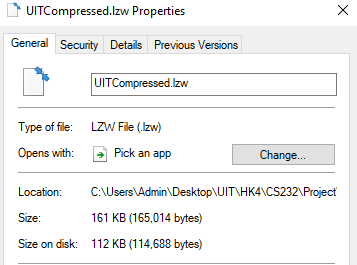
1. **Thực nghiệm trên hình ảnh**

* ***Thực nghiệm 1:*** Ảnh trắng đen (Binary Image)

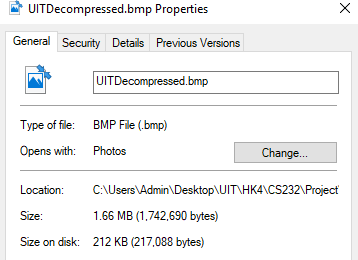
INPUT: ảnh kích thước 1.66 MB

****

**OUTPUT: file nén (.lzw) kích thước 161 KB**

****

**Ảnh đã giải nén: kích thước 1.66 MB**

**** ****

* **Thực nghiệm 2:** Ảnh có màu đơn giản vẽ bằng pixel.

**INPUT**: 3 ảnh pixel có kích thước lần lượt là: 1.37MB, 732KB, 46.8MB.







**OUTPUT**: 3 file nén (.lzw) có kích thước lần lượt là: 188KB, 167KB, 5.62MB.

* ***Thực nghiệm 3:*** Ảnh chụp thực tế, có độ phức tạp về màu sắc.

INPUT: Ảnh có kích thước 10.5 MB.



OUTPUT: file nén (.lzw) có kích thước 22.1 MB.

Tỉ số nén: 2.1 => File nén có kích thước lớn hơn file gốc.

1. **KẾT LUẬN**

**Ưu điểm:**

* Đơn giản, dễ cài đặt, chạy nhanh.
* Là kỹ thuật nén không mất mát.
* Đạt hiệu quả cao với những văn bản tiếng Anh, vì sự lặp lại các ký tự, các chuỗi là thường xuyên.
* Không chỉ văn bản mà LZW còn đạt hiệu quả cao với hình ảnh trắng đen, ảnh pixel có ít độ phức tạp màu.
* Không cần kèm thêm từ điển vào dữ liệu đã nén.

**Khuyết điểm:**

* Với những thông tin không có sự lặp lại thì không thể nén được, thậm chí file nén còn có kích thước lớn hơn file gốc.
* Chỉ đạt hiệu quả cao với dữ liệu dạng văn bản, hình ảnh trắng đen, ảnh pixel đơn giản, còn đối với ảnh chụp thực tế với độ phức tạp cao thì không được như thế, thậm chí không nén được.
* Từ điển có thể chiếm nhiều tài nguyên vì sau rất nhiều lần xét nó có thể tăng kích thước rất nhanh.

1. **TÀI LIỆU THAM KHẢO**
2. *LZW (Lempel–Ziv–Welch) Compression technique -* [*https://www.geeksforgeeks.org/lzw-lempel-ziv-welch-compression-technique/*](https://www.geeksforgeeks.org/lzw-lempel-ziv-welch-compression-technique/)

*2. Lempel–Ziv–Welch -* [*Lempel–Ziv–Welch – Wikipedia*](https://en.wikipedia.org/wiki/Lempel%E2%80%93Ziv%E2%80%93Welch)

*3. LZW COMPRESSION AND DECOMPRESSION -* [*Abstract.pdf (indstate.edu)*](http://cs.indstate.edu/~ngandepalli/Abstract.pdf)

1. **TÀI LIỆU ĐÍNH KÈM**
2. **LZW-Text-Compression** | Công cụ nén và giải nén văn bản bằng thuật toán LZW.
3. **LZW-Image-Compression** | Công cụ nén và giải nén hình ảnh bằng thuật toán LZW.
4. **File PDF**: “LZW Compression.pdf”.
5. **File PowerPoint**: “LZW Compression.pptx”.

**HẾT**