LZW (lempel – Ziv Welch) là phương pháp nén không mất mát thông tin.

1. Ý tưởng:

* Thuật toán LZW sử dụng phép thay thế các ký tự trong chuỗi đầu vào dựa trên bảng mã hóa(giải mã) có sẵn.
* Bảng mã hóa(giải mã) này không cần được lưu kèm với dữ liệu trong quá trình nén. Người mã hóa và người giải mã sẽ tự xây dựng bảng giải mã.
* Bảng mã hóa sẽ được xây dựng giữa vào việc tách chuỗi cần mã hóa hóa thành các chuỗi con theo thứ tự từ điển.

1. Các bước nén dữ liệu theo thuật toán LZW:

* Xây dựng từ điển từ chuỗi đầu vào, đánh số cho các ký tự trong từ điển: chia nhỏ chuỗi cần mã hóa thành nhiều chuỗi con theo thứ tự từ điển của chuỗi con.
* Đánh số thứ tự i cho các chuỗi con.
* Mã hóa từng chuỗi con theo phép thay thế. Mỗi chuỗi con sẽ được thay thế bởi
  + Phần con trỏ, trỏ đến số thứ tự của chuỗi tiền tố: số thứ tự ở dạng nhị phân của chuỗi tiền tố. Phần bit sai khác với chuỗi tiền tố.
* Ghép các phần đã mã hóa của chuỗi con theo thứ tự ban đầu.

1. Các bước giải nén theo thuật toán LZW:

* Dựa vào chuỗi đã được mã hóa, xây dựng lại từ điển.
* Giải mã chuỗi đã mã hóa dựa vào từ điển tự xây dựng lại.

1. Ví dụ về việc mã hóa và giải mã theo thuật toán LZ78:

* Chuỗi cần mã hóa: 1 0110 1010 0010
* Mã hóa
  + Xây dựng các chuỗi con theo thứ thự từ điển. Các chuỗi con lần lượt là: 1, 0, 11, 01, 010, 00, 10
  + Xây dựng bảng mã hóa.
* Trong thuật toán LZ78, chuỗi con thứ n, ta sẽ cần (log2n) bits cho phần con trỏ.
* Phần tử đầu tiên của từ điển có số thự tự 0, và là phần tử rỗng.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Số thứ tự | Số thự tự ở dạng nhị phân | Số bit cần dùng cho con trỏ | Chuỗi con | Chuỗi mã hóa  (con trỏ, bit sai khác) |
| 0 | 000 | 0 | Rỗng (ko bít ký tự) |  |
| 1 | 001 | 1 | 1 | ,1 |
| 2 | 010 | 1 | 0 | 0,0 |
| 3 | 011 | 2 | 11 | 01,1 |
| 4 | 100 | 2 | 01 | 10,1 |
| 5 | 101 | 3 | 010 | 100,0 |
| 6 | 110 | 3 | 00 | 010,0 |
| 7 | 111 | 3 | 10 | 001,0 |

* + Chuỗi mã hóa là: 1 0001 1101 1000 0100 0010
* Giải mã chuỗi : 1 0001 1101 1000 0100 0010
  + Xây dựng lại từ điển: tách chuỗi mã hóa thành các chuỗi con gồm 2 thành phần
    - Phần con trỏ đến số thứ tự chuỗi tiền tố: (log2n) bits
    - Phần bit khác với chuỗi tiền tố.

|  |  |  |  |
| --- | --- | --- | --- |
| Số thứ tự | Số bit cần dùng cho con trỏ | Các chuỗi con  (con trỏ, bit sai khác) | Chuỗi con |
| 1 | 0 | ,1 | 1 |
| 2 | 1 | 0,0 | 0 |
| 3 | 1 | 01,1 | 11 |
| 4 | 2 | 10,1 | 01 |
| 5 | 3 | 100,0 | 010 |
| 6 | 3 | 010,0 | 00 |
| 7 | 3 | 001,0 | 10 |

* + Chuỗi ban đầu là: 1 0110 1010 0010

1. Đặc điểm;

* Không cần phải lưu trữ từ điển, bên giải mã có thể tự xây dựng lại trong quá trình giải mã.
* Số lượng từ trong từ điển tăng lên không giới hạn.
* Các chuỗi con có độ dài lớn sẽ xuất hiện trễ quá trình xây dựng từ điển.
* Ông thêm vào nữa nhé..