Câu 3 Lempel-Ziv-Welch LZW

Công nghệ phần mềm (Vietnam Aviation Academy)

Thuật toán Lempel–Ziv–Welch (LZW)

1) Ý tưởng chính

LZW là một thuật toán nén không mất dữ liệu (lossless) thuộc họ LZ. Thuật toán sử dụng một từ điển để mã hóa các chuỗi lặp lại trong dữ liệu, giúp giảm kích thước tệp.

- Khởi tạo một từ điển chứa tất cả các ký tự đơn (alphabet).
- Duyệt dữ liệu, kết hợp dần các chuỗi mới không có trong từ điển và cấp mã cho chúng.
- Ghi ra mã của các chuỗi đã biết thay vì ghi ký tự thô — tạo hiệu quả nén.

2) Thuật toán mã hóa (encoding)

1. Khởi tạo dictionary với mọi ký tự đơn.
2. w = ""
3. Với mỗi ký tự c trong input:
   - Nếu w + c có trong dictionary: w = w + c
   - Ngược lại: ghi mã của w, thêm w + c vào dictionary, rồi gán w = c
4. Sau khi đọc hết input: ghi mã của w nếu w khác rỗng.

Thực tế thường dùng mã ASCII 0–255 cho các ký tự ban đầu, và code mới bắt đầu từ 256.

3) Ví dụ minh họa

Input: ABAABABA

Khởi tạo: A→1, B→2

Bước	**Ký tự c**	**w trước**	**w+c**	**Trong dict?**	**Hành động**	**Output**	**Thêm vào dict**
1	A	(rỗng)	A	có	w = A	-	-
2	B	A	AB	không	output A=1, add AB→3, w=B	1	AB→3
3	A	B	BA	không	output B=2,	2	BA→4

					add BA→4, w=A		
4	A	A	AA	không	output A=1, add AA→5, w=A	1	AA→5
5	B	A	AB	có (3)	w=AB	-	-
6	A	AB	ABA	không	output AB=3, add ABA→6, w=A	3	ABA→6
7	B	A	AB	có (3)	w=AB	-	-
8	A	AB	ABA	có (6)	w=ABA	-	-
End	-	ABA	-	-	output ABA=6	6	-

Output: 1, 2, 1, 3, 6

Từ điển cuối: {1:A, 2:B, 3:AB, 4:BA, 5:AA, 6:ABA}

4) Giải mã (decoding)

- Khởi tạo dictionary giống encoder.
- Đọc mã đầu → xuất chuỗi tương ứng → gán w = chuỗi đó.
- Với mỗi mã kế tiếp k:
  + Nếu k có trong dictionary: entry = dict[k]
  + Nếu k chưa có: entry = w + ký_tự_đầu(w)
  + Xuất entry, thêm w + ký_tự_đầu(entry) vào dictionary, gán w = entry.
- Kết quả khôi phục được chuỗi gốc.

Ví dụ với output 1,2,1,3,6 → khôi phục được: ABAABABA.

5) Đặc điểm, ưu điểm và nhược điểm

- Ưu điểm: đơn giản, nhanh, không cần truyền từ điển; hiệu quả với dữ liệu có chuỗi lặp.

- Nhược điểm: dữ liệu ngẫu nhiên sẽ nén kém; cần giới hạn hoặc làm mới từ điển.