Abstract

Compression of large collections of data can lead to improvements in retrieval times by offsetting the CPU decompression costs with the cost of seeking and retrieving data from disk. In this paper, the author has study the different compression method which can compress the large DNA sequence. In this paper, authors have explored the DNA compression method that is COMRAD, which is used to compare with the dictionary based compression method i.e. LZ77, LZ78, LZW and general purpose compression method RAY. In this, authors have analyzed which one algorithm is better to compress the large collection of the DNA Sequence. Compression table and the line graph show that which compression algorithm has a better compression ratio and the compression size. It also shows that which one has better compression and decompression time.

References

- D. Wheeler et al., "The Complete Genome of an Individual by Massively Parallel
Survey of Compression of DNA Sequence

- L. Felician and A. Gentili, A nearly optimal Huffman technique in the microcomputer.

**Index Terms**

Computer Science

Artificial Intelligence

**Keywords**

LZ77  LZ78  LZW  RAY  COMRA  DNA Sequence