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 studied the different compression methods 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 methods 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
- Mamta Sharma, "Compression Using Huffman Coding;", IJCSNS International Journal of Computer Science and Network Security, VOL. 10 No. 5, May 2010

Index Terms

Computer Science       Artificial Intelligence

Keywords
LZ77   LZ78   LZW   RAY   COMRA   DNA Sequence