Abstract

Organizations such as IT industry, colleges and Scientists regularly encounter problems to handle large data sets for their different purpose in many areas as for example biological research. These limitations also affect internet search to fetch data, business for analysis etc. So it is simply needed generalized but special types of compression algorithm for dissimilar data to get utmost saving percentage. In this article Compression of biological data that is single and double strand DNA and single strand RNA have been considered. Since biological data are less random compare to any text data that means redundancy within the sequences are more but they have some special property as for example different types of repeat one of such repeat is called dinucleotide repeat. This type of repeat are more in any sequence. Here the two proposed algorithm are based on this repeat using static fixed length LUT for input file and output file mapping.
- R. K. Bharti and Prof. R. K. Singh, &quot;A Biological Sequence Compression based on Look up Table (LUT) using Complementary Palindrome of Fixed Size&quot;, ICJA (0975-8887), Volume 35- No. 11, December 2011.
- Department of Chemistry, Queen Mary University of London, &quot;Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences&quot;.

**Index Terms**

Computer Science

Applied Sciences

**Keywords**

Completely and incompletely specified nucleic acid bases  static LUT  dinucleotide repeats  base pair  sequence line length
compressed sequence length

compression factor

saving percentage