

Data Compression

Enrico Marchionni

`enrico.marchionni@studio.unibo.it`

December 10, 2024

Abstract

Data compression is intended as the practice of reducing the size of binary digital data. It could be considered as a procedure that takes a bit-stream in input and returns another bit-stream as output. The output stream may be of equal length or shorter than the input.

The key to understand data compression is to discuss the distinction between data and information. It can be said that data is how information is represented¹. In simple terms, data can be compressed because its original representation is not the shortest possible. The goal of data compression is to reduce data by maintaining the same information.

The counterpart is that in our time data is intrinsically redundant. And this redundancy is needed. So data compression isn't only a procedure that goes from a bit-stream to another one not longer, but it requires also another procedure that regenerates the original bit-stream of data, necessary for practical use, from the previously given output bit-stream of information.

...

¹ex. the number 0 can be expressed in binary as a sequence of a certain number of zeros, from 1 to ∞ , and we know that calculators use at least 8 bits, let's say n (considering it as a multiple of 8), to represent an integer number. So at the end $n - 1$ bits are redundant in the 0 representation on a calculator.

Contents

...