CLAIMS

1. A data compression apparatus comprising:

a memory; and

a processor that:

extracts a second character string as a matching string from a character string after a first character string in a character string of data before compression that is stored in the memory, the second character string being identical with the first character string, and identifies a length of the matching string, and a relative position indicating how many addresses the first character string precedes the second character string by;

extracts a third character string having a length that is less than the relative position from the extracted second character string;

encodes a length of the third character string; and

encodes the relative position.

2. The data compression apparatus according to claim 1, wherein

when the relative position is less than a prescribed value, the processor extracts the third character string from the second character string.

3. The data compression apparatus according to claim 1, wherein

the processor further performs a process that includes:

extracting, as a non-matching character string, a character string that does not include the second character string identical with the first character string after the first character string in the character string of the data before compression, or a character string that the second character string having a length that is less than a prescribed length; and

extracting, when the length of the third character string is less than a prescribed value, as the non-matching string or a portion of the non-matching string, a maximum value that is a multiple of the length of the third character string, and that does not exceed the prescribed value.

4. The data compression apparatus according to claim 1, wherein the processor defines, as a carton, a minimum unit of recurrence in a recurring character string included in the matching string, and divides the second character string at some or all of positions that each correspond to an integral multiple of the carton from a head of the second character string.

5. The data compression apparatus according to claim 4, wherein the processor divides the second character string at some or all of positions that each correspond to 2^m-1 (where m is an integer) with the carton as a unit from the head of the second character string.

6. The data compression apparatus according to claim 4, wherein when $n=(length+carton-1)\div carton$ is established, where length represents a matching length, and carton represents a length of the carton, the processor divides the second character string at some or all of positions that each correspond to $n\div 2^m$ (where m is an integer) with the carton as a unit from the second character string.

7. A data compression method performed by a computer, the data compression method comprising:

extracting a second character string as a matching string from a character string after a first character string in a character string of data before compression that is stored in a memory, the second character string being identical with the first character string, and identifying a length of the matching string, and a relative position indicating how many addresses the first character string precedes the second character string by;

extracting a third character string having a length that is less than the relative position from the extracted second character string;

encoding a length of the third character string; and

encoding the relative position.

8. A computer-readable recording medium having stored therein a program for causing a computer to execute a process comprising:

extracting a second character string as a matching string from a character string after a first character string in a character string of data before compression that is stored in a memory, the second character string being identical with the first character string, and identifying a length of the matching string, and a relative position indicating how many addresses the first character string precedes the second character string by;

extracting a third character string having a length that is less than the relative position from the extracted second character string;

detecting and encoding a length of the third character string; and

encoding the relative position.