INTERNATIONAL STANDARD

ISO/IEC
12042

Information technology — Data compression for information interchange — Binary arithmetic coding algorithm

Technologies de l'information — Compression de données pour l'échange d'information — Algorithme de codage arithmétique binaire

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 12042:1993
https://standards.iteh.ai/catalog/standards/sist/b53fec01-51c5-4561-9863-
68b1f1883f0b/iso-iec-12042-1993

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 12042:1993
https://standards.iteh.ai/catalog/standards/sist/b53fec01-51c5-4561-9863-
68b1f1883f0b/iso-iec-12042-1993

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 12042 was prepared by the European Computer Manufacturers Association (as Standard ECMA-159) and was adopted, under a special "fast-track procedure", by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, in parallel with its approval by national bodies of ISO and IEC.

Annex A of this International Standard is for information only.

Introduction

In the past decades numerous International Standards for magnetic tapes, magnetic tape cassettes and cartridges, as well as for optical disk cartridges have been published. Media developed recently have a very high physical recording density. In order to make an optimal use of the resulting data capacity, compression algorithms have been designed which allow a reduction of the number of bits required for the representation of user data in coded form.

These compression algorithms are registered by an International Registration Authority set up by ISO/IEC. The registration will consist in allocating to each registered algorithm a numerical identifier which will be recorded on the medium and, thus, indicate which compression algorithm(s) has been used.

The first International Standard for compression algorithms was:

ISO/IEC 11558, *Information technology -Data Compression for Information Interchange - Adaptive Coding with Embedded Dictionary - DCLZ Algorithm*

This International Standard is the next one of this series.

Information technology — Data compression for information interchange — Binary arithmetic coding algorithm

1   Scope

This International Standard specifies an algorithm for the reduction of the number of bits required to represent information. This process is known as data compression. The algorithm uses binary arithmetic coding. The algorithm provides lossless compression and is intended for use in information interchange.

2   Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 11576:1993, Information technology - Procedure for the registration of algorithms for the lossless compression of data.

International Register of Algorithms for Lossless Compression of Data, in accordance with ISO/IEC 11576.

3   Conformance

A compression algorithm shall be in conformance with this International Standard if it satisfies all mandatory requirements of this International Standard.

4   Conventions and notations

The following conventions and notations apply in this International Standard unless otherwise stated:

–   In each field the bytes shall be arranged with Byte 1, the most significant, first. Within each byte the bits shall be arranged with Bit 1, the most significant bit, first and Bit 8, the least significant bit, last.

–   Letters and digits in parentheses represent numbers in hexadecimal notation.

–   The setting of bits is denoted by ZERO or ONE.

–   Numbers in binary notation and bit combinations are represented by strings of ZEROs and ONEs.

–   Numbers in binary notation and bit combinations are shown with the most significant bit to the left.

5   Algorithm identifier

The numeric identifier of this algorithm in the International Register shall be 16.

6   Definitions

For the purposes of this International Standard, the following definitions apply.

6.1 block: A portion of the Logical Data Record, usually having a length of 512 bytes (see 8.2).

6.2 Code Block: A block after compression with a trailer appended.

6.3 Code String: The encoded Logical Data Record.

6.4 encoding: The process of generating Code Blocks from blocks.

6.5 input event: The sample of the input to an encoder currently being examined; in Run Mode it is a byte; in Normal Mode it is a bit.

6.6 Logical Data Record: The data entity that is the input to the data compressor.

6.7 trailer: data appended to a block after compression and addition of pad bits.

6.8 Unique Table Pair: The last of the 256 Table Pairs, used only in Run Mode.

7 List of acronyms

CV   Current Value
EV   Estimated Value
LDR   Logical Data Record
TP   Table Pair

8 Compression algorithm

8.1 General

The LDR is transformed to a Code String by a one-pass, adaptive encoding technique designed to provide lossless data compression. By the use of a suitable decoding technique the exact original LDR can be recovered from the Code String.

8.2 Encoders

The LDR shall be divided into 512-byte blocks, except for the last block, which may be of any length less than, or equal to, 512 bytes. The blocks shall be routed sequentially to eight encoders, numbered from 0 to 7, commencing with encoder 0. If the LDR contains more than 4 096 bytes the compressor shall return to encoder 0 and repeat the process (see figure 2).

8.3 Formation of a Code Block

The output of each encoder is a Code Block (see figure 1).

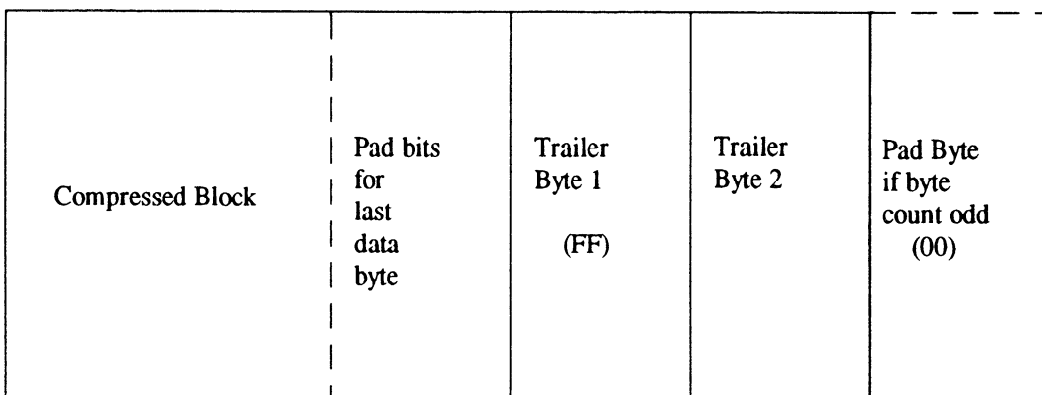| Compressed Block	Pad bits for last data byte	Trailer Byte 1 (FF)	Trailer Byte 2	Pad Byte if byte count odd (00)

Figure 1 - Code Block

Since the degree of compression achieved in an encoder depends upon the relative frequency of the bit patterns in the LDR, and upon the presence of sequences of identical bytes, the length of the Compressed Block cannot be predicted. Pad bits set to ZERO shall be added at the end to form an integral number of 8-bit bytes.

The Code Block shall be completed by appending a trailer. The trailer shall consist of two Trailer Bytes possibly followed by a Pad Byte.

Trailer Byte 1   shall be set to (FF).

Trailer Byte 2

Bits 1 to 4   shall be set to 1100 if the Code Block has been generated from the last block of the LDR.
shall be 1001 for all other Code Blocks.

Bit 5   shall be set to ZERO if the number of bytes after encoding is even.
shall be set to ONE if the number of bytes after encoding is odd.

Bits 6 to 8   shall specify the number of pad bits that have been added to form an integral number of bytes.

If the number of bytes in the Compressed Block plus the pad bits, is odd, a Pad Byte set to (00) shall be appended after Trailer Byte 2 to give an even number of bytes.

8.4   Code String

The Code String shall be assembled from the outputs of the encoders, with the first portion being that generated by encoder 0, the second that generated by encoder 1, and so on (see figure 2).

8.5   Table Pairs

Each encoder shall be allocated a table of 256 pairs of numbers, numbered from 1 to 256. The first number of each Table Pair shall be the estimated value (EV) of the Input Event to be encoded; it shall be 1 or 0. The second number (K) shall be a measure of the probability of the Input Event being equal to the EV. K shall have the value 1, 2, 3 or 4, with the probability shown in table 1.

Table 1- Probability values of K

K	Probability
1	1 - 2
2	2 - 4
3	4 - 8
4	8 - 16

The probabilities shall be a measure of how much more likely it is that the value of the Input Event is equal to the EV rather than being unequal (e.g. for $K=2$ the probability that the Input Event is equal to the EV shall be 2 to 4 times as great as the probability that it would not be equal).

Before commencing the encoding of the LDR all EVs shall be set to ZERO and all values of K shall be set to ONE.

8.6   Encoding

The data shall first be examined on a byte basis. Bytes shall be fetched sequentially from the block, starting with the first byte, and compared with the previous byte. The first byte in a block shall be compared with (40).

Run Mode (see 8.6.2) shall be disabled when the first byte is fetched.

If the current byte differs from the previous byte and Run Mode is not enabled, then the byte shall be encoded, bit by bit, in Normal Mode (see 8.6.1).

If the current byte differs from the previous byte and Run Mode is enabled, then encoding shall proceed as defined in 8.6.2.2.

If the two bytes are identical and Run Mode is not enabled, then Run Mode shall be enabled and the byte shall be encoded, bit by bit, in Normal Mode.

If the two bytes are identical and Run Mode is enabled, then encoding shall proceed as defined in 8.6.2.1.

8.6.1 Normal Mode

The first (most significant) bit of the byte shall be compared with the EV in the first Table Pair. Depending upon the result of this comparison, one of two actions, which are described in 8.6.1.1, shall result. The choice of which Table Pair to select for the remaining bits of the byte shall be determined by the bits previously encoded in the byte.

The first bit of each byte shall always use the first Table Pair.

The second bit shall use the second or third Table Pair, depending upon whether the first bit was ZERO or ONE, respectively.

The third bit shall use one of the next four Table Pairs depending upon the first two bits. The fourth Table Pair shall be used if the first two bits were 00, and so on (see figure 3).

This procedure requires the first 255 Table Pairs. The remaining Table Pair is the Unique Table Pair; it shall be used in the Run Mode only (see 8.6.2).

The process of encoding is then performed in two stages:

- The bit is encoded as in 8.6.1.1.

- The values of K and EV are revised as described in 8.6.1.2.

8.6.1.1 Bit encoding

Two values shall be developed during the bit comparison portion of the compression process. Both are fractional binary numbers to four binary places. One value, termed the Current Value (CV), shall be used to generate the output (Code Block).

The second value, termed the Width, shall have values in the range 0,0000 to 1,1111.

For each block the CV shall be initialized to 0,0000 and the Width value shall be initialized to 1,0000.

Each Input Event shall cause the two values to be modified according to the following rules:

i)   The Input Event is equal to the EV

The CV shall be increased by 2^{-K}

The Width shall be decreased by 2^{-K}

where K is the measure of the probability of the Input Event (see 8.5).

If the Code Block is not null the bit to the left of the point in the CV shall be logically added to the last bit appended to the Code Block and replaced in the CV by a ZERO. If this addition causes the most recently generated complete byte in the Code Block to become (FF), then (0) shall be appended to the Code Block at the end of that byte to prevent a carry from propagating beyond the last complete byte.

If the Code Block is null the bit to the left of the point in the CV will already be a ZERO.

The Width shall then be compared with 1,0000.

If it is equal to, or greater than, 1,0000, then the Table Pair shall be revised (see 8.6.1.2) and the next bit fetched for encoding.

If it is less than 1,0000, then all the bits of the Width shall be shifted left one position; the leftmost bit, set to ZERO, shall be dropped, and the rightmost position shall be filled with a ZERO. The bit in the first position to the right of the point in the CV shall be appended to the encoder output as part of the Code Block. The remaining bits to the right of the point in the CV shall be shifted left one position and the rightmost position shall be filled with a ZERO.

To prevent a carry from propagating beyond the last complete byte, each byte in the Code Block shall be examined as it is completed. If it equals (FF) then (0) shall be appended to the Code Block.

ii)   The Input Event is not equal to the EV.

The Width shall be reset to 1,0000.

The *K* bits to the right of the point in the CV shall be shifted out and appended to the Code Block, the leftmost bit first. As each bit is appended to the Code Block a check shall be made to determine whether a byte boundary has been reached. If it has, then the byte just completed shall be checked; if it equals (FF) then (0) shall be appended to the Code Block.

The rightmost *K* bit positions of the CV shall be set to ZEROs.

8.6.1.2   Revision of *K* and EV (see figure 4)

As each Input Event is compared with the EV the values of *K* and the EV for that Table Pair shall be amended.

A four-stage binary counter (Mc counter) shall be set to 0000 at the beginning of each block and incremented as described below.

i)   The Input Event is equal to the EV

K shall be incremented according to table 2, where bits marked X shall be ignored.

Table 2 - Rules for incrementing *K*

Current value of *K*	State of Counter	New value of *K*
1	XX11	2
2	X111	3
3	1111	4
4	XXXX	4

For all other states of the counter *K* shall be unchanged.

The counter shall then be incremented by 1. If the counter value was 1111 incrementing by 1 shall result in a counter value of 0000.

The EV shall be unchanged.

ii)   The Input Event is not equal to the EV

For *K* greater than 1   :   The value of *K* shall be decremented by 1

the counter shall not be incremented

the EV shall be unchanged

For *K* equal to 1   :   the value of *K* shall be unchanged

the counter shall not be incremented

the EV shall be inverted

8.6.2   Run Mode

Run Mode shall have been enabled if the last two bytes to be encoded were identical.

8.6.2.1   If the current byte is identical with these two bytes, then the Unique Table Pair is selected and the EV is compared with ONE; the values of the Width, CV, *K* and EV are amended, as defined in 8.6.1.1 and 8.6.1.2. This may result in the Code Block being extended.