INTERNATIONAL STANDARD

ISO/IEC 15438

Third edition
2015-09-15

Information technology — Automatic identification and data capture techniques — PDF417 bar code symbology specification

Technologies de l'information — Techniques automatiques d'identification et de capture des données — Spécifications pour la symbologie de code à barres PDF417

Reference number
ISO/IEC 15438:2015(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 31, *Automatic identification and data capture techniques*.

This third edition cancels and replaces the second edition (ISO/IEC 15438:2006), of which it constitutes a minor revision.

Introduction

The technology of bar coding is based on the recognition of patterns of bars and spaces of defined dimensions. There are various methods of encoding information in bar code form, known as symbologies, and the rules defining the translation of characters into bars and space patterns and other essential features are known as the symbology specification.

Manufacturers of bar code equipment and users of bar code technology require publicly available standard symbology specifications to which they can refer when developing equipment and application standards. It is the intent and understanding of ISO/IEC that the symbology presented in this International Standard is entirely in the public domain and free of all user restrictions, licences and fees.

Information technology — Automatic identification and data capture techniques — PDF417 bar code symbology specification

1 Scope

This International Standard specifies the requirements for the bar code symbology known as PDF417. It specifies PDF417 symbology characteristics, data character encodation, symbol formats, dimensions, error correction rules, reference decoding algorithm, and a number of application parameters.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 646, *Information technology — ISO 7-bit coded character set for information interchange*

ISO/IEC 15415, *Information technology — Automatic identification and data capture techniques — Bar code symbol print quality test specification — Two-dimensional symbols*

ISO/IEC 15424, *Information technology — Automatic identification and data capture techniques — Data Carrier Identifiers (including Symbology Identifiers)*

ISO/IEC 19762-1, *Information technology — Automatic identification and data capture (AIDC) techniques — Harmonized vocabulary — Part 1: General terms relating to AIDC*

ISO/IEC 19762-2, *Information technology — Automatic identification and data capture (AIDC) techniques — Harmonized vocabulary — Part 2: Optically readable media (ORM)*

ISO/IEC 24723, *Information technology — Automatic identification and data capture techniques — GS1 Composite bar code symbology specification*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19762-1, ISO/IEC 19762-2 and the following apply.

3.1
basic channel model
standard system for encoding and transmitting bar code data where data message bytes are output from the decoder but no control information about the message is transmitted

Note 1 to entry: A decoder complying with this model operates in Basic Channel Mode.

3.2
bar-space sequence
sequence which represents the module widths of the elements of a symbol character

3.3
cluster
any of the three mutually exclusive subsets of PDF417 symbol characters

Note 1 to entry: The symbol characters in a given cluster conform with particular structural rules which are used in decoding the symbology.

3.4
compaction mode
any of the three data compaction algorithms in PDF417 (Text, Numeric and Byte Compaction modes) which are used to map 8-bit data bytes efficiently to PDF417 codewords

3.5
e-distance
distance from the leading edge of an element to the leading edge of the next similar element, or from trailing edge to trailing edge

3.6
error correction codeword
encodes a value derived from the error correction codeword algorithm to enable decode errors to be detected and, depending on the error correction level, to be corrected

3.7
Extended Channel Interpretation
ECI
procedure within some symbologies, including PDF417, to replace the default interpretation with another interpretation in a reliable manner

Note 1 to entry: The interpretation intended prior to producing the symbol can be retrieved after decoding the scanned symbol to recreate the data message in its original format.

3.8
Extended Channel Model
system for encoding and transmitting both data message bytes and control information about the message, the control information being communicated using Extended Channel Interpretation (ECI) escape sequences

Note 1 to entry: A decoder complying with this model operates in Extended Channel Mode.

3.9
function codeword
initiates a particular operation within a symbology

EXAMPLE To switch between data encoding sets, to invoke a compaction scheme, to program the reader, or to invoke Extended Channel Interpretations.

3.10
Global Label Identifier
GLI
procedure in the PDF417 symbology which behaves in a similar manner to Extended Channel Interpretation

Note 1 to entry: The GLI system was the PDF417-specific precursor to the symbology-independent ECI system.

3.11
Macro PDF417
procedure in the PDF417 symbology logically to distribute data from a computer file across a number of related PDF417 symbols

Note 1 to entry: The procedure considerably extends the data capacity beyond that of a single symbol.

Note 2 to entry: This procedure is similar to the Structured Append feature in other symbologies.

3.12
Mode Latch codeword
used to switch from one mode to another mode, which stays in effect until another latch or shift codeword is implicitly or explicitly brought into use, or until the end of the symbol is reached

3.13
Mode Shift codeword
used to switch from one mode to another for one codeword, after which encoding returns to the original mode

3.14
Row Indicator codeword
PDF417 codeword adjacent to the start or stop character in a row, which encodes information about the structure of the PDF417 symbol in terms of the row identification, total number of rows and columns, and the error correction level

3.15
Symbol Length Descriptor
first codeword in a PDF417 symbol, which encodes the total number of data codewords in the symbol

4 Symbols, operations and abbreviated terms

4.1 Symbols

For the purposes of this International Standard, the following mathematical symbols apply. There are some cases where the symbols below have been used in a different manner in an equation. This has been done for consistency with a more general use of the notation and is always clearly defined in the text.

A symbol aspect ratio (height to width) of a PDF417 symbol

b element width in a symbol character

c number of columns in the symbol in the data region (excluding start, stop and row indicator codewords)

d data codeword including all function codewords

E error correction codeword

e edge to similar edge dimension in a symbol character

F row number

f number of substitution errors

H height of symbol including quiet zone

K cluster number

k number of error correction codewords

L left row indicator

l number of erasures

m number of source data codewords prior to the addition of the Symbol Length Descriptor and any pad codewords

n total number of data codewords including Symbol Length Descriptor and any pad codewords

p pitch or width of a symbol character

Q_H horizontal quiet zone

Q_V vertical quiet zone

R right row indicator

r number of rows in the symbol

s error correction level

W width of symbol including quiet zone

X X-dimension or module width

Y module height (also called row height)

4.2 Mathematical operations

For the purposes of this International Standard, the following mathematical operations apply.

div is the integer division operator, rounding down

INT is the integer value, i.e. where a number is rounded down to its whole number component, ignoring its decimal fractions

mod is the positive integer remainder after division. If the remainder is negative, add the value of the divisor to make it positive. For example, the remainder of –29 160 divided by 929 is –361 which when added to 929 yields 568.

4.3 Abbreviated terms

For the purposes of this International Standard, the following abbreviated terms apply.

ECI Extended Channel Interpretation

GLI Global Label Identifier

5 Requirements

5.1 Symbology characteristics

5.1.1 Basic characteristics

PDF417 is a bar code symbology with the following basic characteristics.

a) Encodable character set:

1) Text Compaction mode (see 5.4.1.5) permits all printable ASCII characters to be encoded, i.e. values 32 to 126 inclusive in accordance with ISO/IEC 646 (IRV), as well as selected control characters;

2) Byte Compaction mode (see 5.4.3) permits all 256 possible 8-bit byte values to be encoded. This includes all ASCII characters value 0 to 127 inclusive and provides for international character set support;

3) Numeric Compaction mode (see 5.4.4) permits efficient encoding of numeric data strings;

4) Up to 811 800 different character sets or data interpretations;

5) Various function codewords for control purposes.

b) Symbol character structure: (*n, k, m*) characters of 17 modules (*n*), 4 bar and 4 space elements (*k*), with the largest element 6 modules wide (*m*).

c) Maximum possible number of data characters per symbol (at error correction level 0): 925 data codewords which can encode:

1) Text Compaction mode: 1 850 characters (at 2 data characters per codeword);

2) Byte Compaction mode: 1 108 characters (at 1,2 data characters per codeword);

3) Numeric Compaction mode: 2 710 characters (at 2,93 data characters per codeword).

At the minimum recommended error correction level, there is a maximum of 863 data codewords which can encode:

4) Text Compaction mode: 1 726 characters (at 2 data characters per codeword);

5) Byte Compaction mode: 1 033 characters (at 1,2 data characters per codeword);

6) Numeric Compaction mode: 2 528 characters (at 2,93 data characters per codeword).

d) Symbol size:

1) Number of rows: 3 to 90;

2) Number of columns: 1 to 30;

3) Width in modules: 90X to 583X including quiet zones;

4) Maximum codeword capacity: 928 codewords;

5) Maximum data codeword capacity: 925 codewords.

Since the number of rows and the number of columns are selectable, the aspect ratio of a PDF417 symbol may be varied when printing to suit the spatial requirements of the application.

e) Selectable error correction: 2 to 512 codewords per symbol (see 5.7).

f) Non-data overhead:

1) Per row: 73 modules, including quiet zones;

2) Per symbol: a minimum of 3 codewords, represented as symbol characters.

g) Code type: continuous, multi-row two-dimensional.

h) Character self-checking: Yes.

i) Bi-directionally decodable: Yes.

5.1.2 Summary of additional features

Additional features which are inherent or optional in PDF417 are summarised below.

a) **Data compaction**: (inherent) Three schemes are defined to compact a number of data characters into codewords. Generally data is not directly represented on a one character for one codeword basis (see 5.4.1.5 to 5.4.4).

b) **Extended Channel Interpretations**: (optional) These mechanisms allow up to 811 800 different data character sets or interpretations to be encoded (see 5.5).

c) **Macro PDF417**: (optional) This mechanism allows files of data to be represented logically and consecutively in a number of PDF417 symbols. Up to 99 999 different PDF417 symbols can be so linked or concatenated and be scanned in any sequence to enable the original data file to be correctly reconstructed (see 5.13).

d) **Edge to edge decodable**: (inherent) PDF417 can be decoded by measuring elements from edge to similar edge (see 5.3.1).

e) **Cross row scanning**: (inherent) The combination of the following three characteristics in PDF417 facilitates cross row scanning:

   1) being synchronised horizontally, or self clocking;

   2) row identification;

   3) being synchronised vertically, by using the cluster values to achieve local row discrimination.

   This combination allows a single linear scan to cross a number of rows and achieve a partial decode of the data so long as at least one complete symbol character per row is decoded into its codeword. The decoding algorithm can then place the individual codewords into a meaningful matrix.

f) **Error correction**: (inherent) A user may define one of 9 error correction levels. All but Level 0 not only detect errors but also can correct erroneously decoded or missing codewords (see 5.7).

g) **Compact PDF417**: (optional) In relatively 'clean' environments, it is possible to reduce some of the row overhead to improve the symbol density (see 5.12).

NOTE In earlier specifications of PDF417, Compact PDF417 was called Truncated PDF417. Compact PDF417 is the preferred term to avoid confusion with the more general use of the term 'truncated'.

5.2 Symbol structure

5.2.1 PDF417 symbol parameters

Each PDF417 symbol consists of a stack of vertically aligned rows with a minimum of 3 rows (maximum 90 rows). Each row shall include a minimum of 1 symbol character (maximum 30 symbol characters), excluding start, stop and row indicator columns. The symbol shall include a quiet zone on all four sides. Figure 1 illustrates a PDF417 symbol encoding the text: PDF417 Symbology Standard.

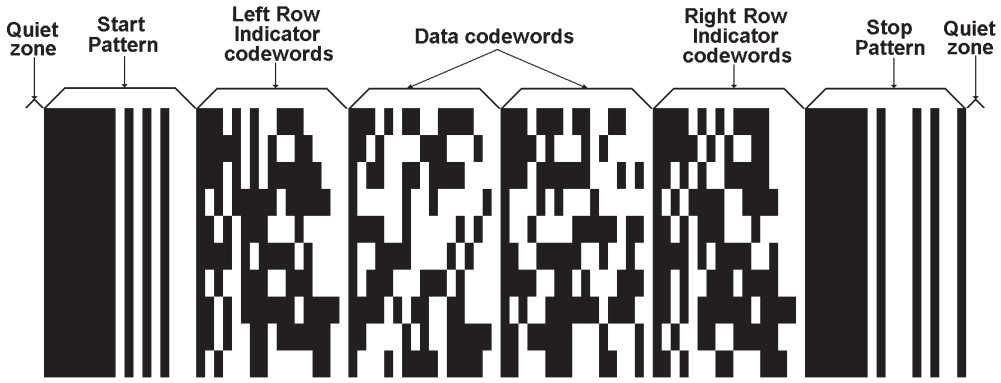

Figure 1 — PDF417 symbol structure

5.2.2 Row parameters

Each PDF417 row shall comprise of the following:

a) leading quiet zone;

b) start character;

c) left row indicator symbol character;

d) 1 to 30 symbol characters;

e) right row indicator symbol character;

f) stop character;

g) trailing quiet zone.

NOTE The number of symbol characters (or codewords) defined in item 'd' above is equal to the number of data columns in the PDF417 symbol.

5.2.3 Codeword sequence

A PDF417 symbol may contain up to 928 symbol characters or codewords. Symbol character is the more appropriate term to refer to the printed bar/space pattern; codeword is more appropriate for the numeric value of the symbol character. The codewords shall follow this sequence:

a) The first codeword, the Symbol Length Descriptor, shall always encode the total number of data codewords in the symbol, including the Symbol Length Descriptor itself, data codewords and pad codewords, but excluding the number of error correction codewords.

b) The data codewords shall follow, from the most significant encodable character. Function codewords may be inserted to achieve data compaction.

c) Pad codewords to enable the codeword sequence to be represented in a rectangular matrix. Pad codewords may also be used to fill additional complete rows to achieve an aspect ratio desired or as specified by the application.

d) An optional Macro PDF417 Control Block.

e) Error correction codewords for error detection and correction.

The codewords are arranged with the most significant codeword adjacent to the Symbol Length Descriptor, and are encoded from left to right and from top row to bottom. Figure 2 illustrates in layout format the sequence for a symbol like what is being shown in Figure 1. In Figure 2, an error correction level of 1 has been used and one pad character was needed to completely fill the symbol matrix.

S T A R T	L_1	d_{15}	d_{14}	R_1	**S T O P**
	L_2	d_{13}	d_{12}	R_2	
	L_3	d_{11}	d_{10}	R_3	
	L_4	d_9	d_8	R_4	
	L_5	d_7	d_6	R_5	
	L_6	d_5	d_4	R_6	
	L_7	d_3	d_2	R_7	
	L_8	d_1	d_0	R_8	
	L_9	E_3	E_2	R_9	
	L_{10}	E_1	E_0	R_{10}	

Figure 2 — PDF417 Example of Symbol Layout Schematic

where

L, R, d and E	are as defined in Clause 4;
d_{15}	Symbol Length Descriptor (in this example, with a value of 16);
d_{14} to d_1	encoded representation of data;
d_0	pad codeword.

The rules and advice for structuring the matrix are included in 5.9.

5.3 Basic encodation

5.3.1 Symbol character structure

Each PDF417 symbol character shall consist of four bar elements and four space elements, each of which can be one to six modules wide. The four bar and four space elements shall measure 17 modules in total. PDF417 symbol characters can be decoded by measuring the e-distances within the character.

Each symbol character is defined by an 8-digit bar-space sequence which represents the module widths of the eight elements of that symbol character. Figure 3 illustrates a symbol character with the bar-space sequence 51111125.

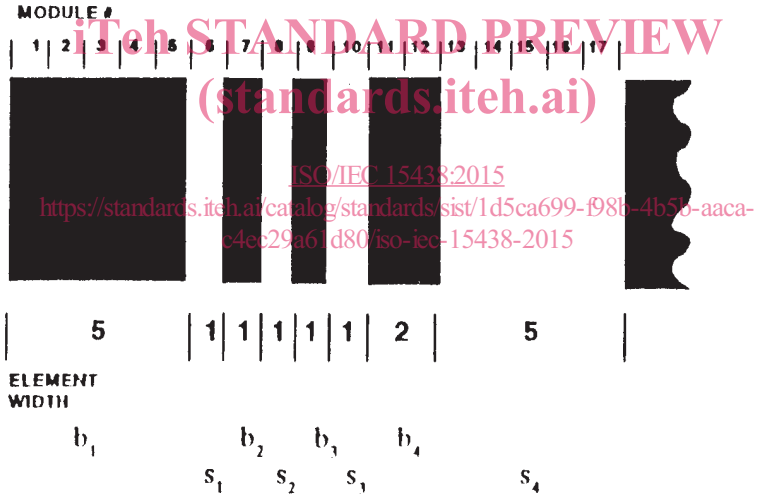

Figure 3 — A PDF417 symbol character

There are 929 defined symbol character values (codewords) numbered from 0 to 928.

The codewords are represented by three mutually exclusive symbol character sets, or clusters. Each cluster encodes the 929 available PDF417 codewords into different bar-space patterns so that one cluster is distinct from another. The cluster numbers are 0, 3, and 6. The cluster definition applies to all PDF417 symbol characters, except for start and stop characters.

The cluster number K is defined by the following formula:

$$K = (b_1 - b_2 + b_3 - b_4 +) \bmod 99$$

where b_1, b_2, b_3 and b_4 represent the width in modules of the four bar elements respectively.

The cluster number *K* for the symbol character in Figure 3 is:

$K = (5 - 1 + 1 - 2 + 9) \bmod 9 = 3$

The codewords and the bar-space sequences for each cluster of symbol characters are given in Annex A.

5.3.2 Start and stop characters

The start and stop characters shall be composed as defined in Table 1 and illustrated in Figure 4:

Table 1 — Bar-space sequence for Start and Stop Characters

Character	Bar-space sequence								
	B	S	B	S	B	S	B	S	B
Start	8	1	1	1	1	1	1	3	
Stop	7	1	1	3	1	1	1	2	1

NOTE 1 The PDF417 stop and start characters are unique in having elements more than 6 modules wide.

NOTE 2 The stop character has one extra single module bar element.

The start and stop characters shall have the same bar-space sequence for all rows.

Figure 4 — PDF417 Start and Stop Characters

5.4 High level (data) encodation

High level encoding converts the data characters into their corresponding codewords.

Data compaction schemes shall be used to achieve efficient high level encoding. Three modes are defined below, each of which defines a particular efficient mapping between user defined data and codeword sequences. PDF417 has three data compaction modes:

— Text Compaction mode (see 5.4.1.5);

— Byte Compaction mode (see 5.4.3);

— Numeric Compaction mode (see 5.4.4).

A given string of data bytes may be represented by different codeword sequences, depending on how the encoder switches between compaction modes and sub-modes. There is no single specified way to encode data in a PDF417 symbol.

900 codewords (0 to 899) are available in each mode for data encodation and other functions within the mode. The remaining 29 codewords are assigned to specific functions (see 5.4.1) independent of the current compaction mode.