

NOTE 4 This is useful when the language cannot be determined from the character codes used in the text.

The escape sequence shall consist of the following elements, in order:

- a) The Unicode value U+001B (that is, the byte sequence 0 followed by 27).
- b) A 2- byte ISO 639 language code.

EXAMPLE en for English or ja for Japanese encoded as ASCII characters.

- c) (*Optional*) A 2-byte ISO 3166 country code.

EXAMPLE US for the United States or JP for Japan.

- d) The Unicode value U+001B.

NOTE 5 The complete list of codes defined by ISO 639 and ISO 3166 can be obtained from the International Organization for Standardization (see the Bibliography).

7.9.2.3 PDFDocEncoded String Type

A PDFDocEncoded string is a character string in which the characters shall be represented in a single byte using PDFDocEncoding.

NOTE PDFDocEncoding does not support all Unicode characters whereas UTF-16BE does.

7.9.2.4 Byte String Type

The byte string type shall be used for binary data that shall be represented as a series of bytes, where each byte may be any value representable in 8 bits. Byte string type is a subtype of string type.

NOTE The string may represent characters but the encoding is not known. The bytes of the string may not represent characters.

7.9.3 Text Streams

A *text stream* (PDF 1.5) shall be a PDF stream object (7.3.8, "Stream Objects") whose unencoded bytes shall meet the same requirements as a text string (7.9.2.2, "Text String Type") with respect to encoding, byte order, and lead bytes.

7.9.4 Dates

Date values used in a PDF shall conform to a standard date format, which closely follows that of the international standard ASN.1 (Abstract Syntax Notation One), defined in ISO/IEC 8824. A date shall be a text string of the form

(D: YYYYMMDDHHmmSSOHH' mm)

where:

YYYY shall be the year

MM shall be the month (01–12)

DD shall be the day (01–31)

HH shall be the hour (00–23)

mm shall be the minute (00–59)

SS shall be the second (00–59)

O shall be the relationship of local time to Universal Time (UT), and shall be denoted by one of the characters PLUS SIGN (U+002B) (+), HYPHEN-MINUS (U+002D) (-), or LATIN CAPITAL LETTER Z (U+005A) (Z) (see below)

HH followed by APOSTROPHE (U+0027) (') shall be the absolute value of the offset from UT in hours (00–23)

mm shall be the absolute value of the offset from UT in minutes (00–59)

The prefix *D:* shall be present, the year field (YYYY) shall be present and all other fields may be present but only if all of their preceding fields are also present. The APOSTROPHE following the hour offset field (HH) shall only be present if the HH field is present. The minute offset field (mm) shall only be present if the APOSTROPHE following the hour offset field (HH) is present. The default values for *MM* and *DD* shall be both 01; all other numerical fields shall default to zero values. A PLUS SIGN as the value of the *O* field signifies that local time is later than UT, a HYPHEN-MINUS signifies that local time is earlier than UT, and the LATIN CAPITAL LETTER Z signifies that local time is equal to UT. If no UT information is specified, the relationship of the specified time to UT shall be considered to be GMT. Regardless of whether the time zone is specified, the rest of the date shall be specified in local time.

EXAMPLE For example, December 23, 1998, at 7:52 PM, U.S. Pacific Standard Time, is represented by the string *D:199812231952-08'00*

7.9.5 Rectangles

Rectangles are used to describe locations on a page and bounding boxes for a variety of objects. A rectangle shall be written as an array of four numbers giving the coordinates of a pair of diagonally opposite corners.

NOTE Although rectangles are conventionally specified by their lower-left and upper-right corners, it is acceptable to specify any two diagonally opposite corners. Applications that process PDF should be prepared to normalize such rectangles in situations where specific corners are required.

Typically, the array takes the form

$[ll_x \ ll_y \ ur_x \ ur_y]$

specifying the lower-left *x*, lower-left *y*, upper-right *x*, and upper-right *y* coordinates of the rectangle, in that order. The other two corners of the rectangle are then assumed to have coordinates (ll_x, ur_y) and (ur_x, ll_y) .

7.9.6 Name Trees

A *name tree* serves a similar purpose to a dictionary—associating keys and values—but by different means. A name tree differs from a dictionary in the following important ways:

- Unlike the keys in a dictionary, which are name objects, those in a name tree are strings.
- The keys are ordered.
- The values associated with the keys may be objects of any type. Stream objects shall be specified by indirect object references (7.3.8, "Stream Objects"). The dictionary, array, and string objects should be specified by indirect object references, and other PDF objects (nulls, numbers, booleans, and names) should be specified as direct objects.
- The data structure can represent an arbitrarily large collection of key-value pairs, which can be looked up efficiently without requiring the entire data structure to be read from the PDF file. (In contrast, a dictionary can be subject to an implementation limit on the number of entries it can contain.)