

NAME

encoding – Manipulate encodings

SYNOPSIS**encoding** *option* ?*arg arg ...*?**INTRODUCTION**

Strings in Tcl are logically a sequence of 16-bit Unicode characters. These strings are represented in memory as a sequence of bytes that may be in one of several encodings: modified UTF-8 (which uses 1 to 3 bytes per character), 16-bit “Unicode” (which uses 2 bytes per character, with an endianness that is dependent on the host architecture), and binary (which uses a single byte per character but only handles a restricted range of characters). Tcl does not guarantee to always use the same encoding for the same string.

Different operating system interfaces or applications may generate strings in other encodings such as Shift-JIS. The **encoding** command helps to bridge the gap between Unicode and these other formats.

DESCRIPTION

Performs one of several encoding related operations, depending on *option*. The legal *options* are:

encoding convertfrom ?*encoding*? *data*

Convert *data* to Unicode from the specified *encoding*. The characters in *data* are treated as binary data where the lower 8-bits of each character is taken as a single byte. The resulting sequence of bytes is treated as a string in the specified *encoding*. If *encoding* is not specified, the current system encoding is used.

encoding convertto ?*encoding*? *string*

Convert *string* from Unicode to the specified *encoding*. The result is a sequence of bytes that represents the converted string. Each byte is stored in the lower 8-bits of a Unicode character (indeed, the resulting string is a binary string as far as Tcl is concerned, at least initially). If *encoding* is not specified, the current system encoding is used.

encoding dirs ?*directoryList*?

Tcl can load encoding data files from the file system that describe additional encodings for it to work with. This command sets the search path for *.enc encoding data files to the list of directories *directoryList*. If *directoryList* is omitted then the command returns the current list of directories that make up the search path. It is an error for *directoryList* to not be a valid list. If, when a search for an encoding data file is happening, an element in *directoryList* does not refer to a readable, searchable directory, that element is ignored.

encoding names

Returns a list containing the names of all of the encodings that are currently available. The encodings “utf-8” and “iso8859-1” are guaranteed to be present in the list.

encoding system ?*encoding*?

Set the system encoding to *encoding*. If *encoding* is omitted then the command returns the current system encoding. The system encoding is used whenever Tcl passes strings to system calls.

EXAMPLE

The following example converts a byte sequence in Japanese euc-jp encoding to a TCL string:

```
set s [encoding convertfrom euc-jp "\xA4\xCF"]
```

The result is the unicode codepoint: “\u306F”, which is the Hiragana letter HA.

SEE ALSO

Tcl_GetEncoding(3), fconfigure(n)

KEYWORDS

encoding, unicode