

**NAME**

`utf8gen` – Generate UTF-8 output from hexadecimal input

**SYNOPSIS**

```
utf8gen [ [-e format1] | [-E format2] ] [-r formatr]
          [ [-u utf8_format] | -n] [-c] [-s]
          [-i input_file] [-o output_file]
```

**DESCRIPTION**

**utf8gen** reads a list of hexadecimal ASCII values in the range 0 through 10FFFF, one per line, and prints the UTF-8 encoding of that number as a Unicode code point.

Each input line must begin with a hexadecimal number. A string may follow after that, which can be echoed to the output as the "remainder" (see the `-r` option below). The total input line length, including an ending newline, is limited to 4096 bytes.

**OPTIONS**

- `-c` After the UTF-8 codes are printed, print a space followed by the character that the hexadecimal code point represents.
- `-e` Echo the input code point in one format, using the `printf(3)` format string *format1*.
- `-E` Echo the input code point in two formats, using the `printf(3)` format string *format2*.
- `-n` Do *not* print the UTF-8 byte values. This can be useful if only the printed character itself is desired; see the `-c` option.
- `-r` Print the remainder of the input string after the initial hexadecimal digits, using the `printf(3)` format string *formatr*.
- `-s` Swap the order of output: print the UTF-8 output portion first, then print the input string portion. This can be useful for generating code containing a UTF-8 encoding followed by a comment that contains the input hexadecimal digits.
- `-u` Print the UTF-8 encoded value of the input hexadecimal number, as numeric codes for each UTF-8 byte, using the `printf(3)` format string *utf8\_format*. If no string is specified, a default format of a backslash followed by three octal digits is printed for each byte.

**EXAMPLES**

```
utf8gen -e "0x%04X " -u "%03o"
utf8gen -E "U+%04x = 0%02o = "
utf8gen -s -e " /* U+%04X */" -u "%03o"
```

**FILES**

Files contain lines that each begin with an ASCII hexadecimal code in the valid Unicode range 0 through 10FFFF, inclusive. This hexadecimal code may optionally be followed by a space followed by an arbitrary string ending with a newline, up to the limit of 4096 bytes per input line. An example line could be the following (with no indent):

```
41 Letter 'A'
```

**SEE ALSO**

For more detailed explanations and examples of common usage, consult the **utf8gen** `texinfo` manual.

**AUTHOR**

**utf8gen** was written by Paul Hardy.

**LICENSE**

**utf8gen** is Copyright © 2018 Paul Hardy.

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

**BUGS**

No known bugs exist.