

NAME

olden – convert modern English letters into archaic forms

SYNOPSIS

olden [*options*] [-i *input_filename*] [-o *output_filename*]

DESCRIPTION

olden(1) reads an input file containing English text and converts letters to archaic letters in UTF-8. The output file can optionally begin with a UTF-8 "Byte Order Mark" (though the notion of byte ordering is an oxymoron with UTF-8, which is independent of machine byte ordering).

OPTIONS**GENERAL INFORMATION**

--version Print program name, version, and other general information to STDOUT, and then exit.

I/O OPTIONS

The following three options affect file input and output.

- b Begin output file with a UTF-8 Byte Order Mark, which is Unicode code point U+FEFF. In UTF-8, this is the three byte octal sequence 357 273 277.
- i Specify the input file. The default is STDIN.
- o Specify the output file. The default is STDOUT.

LETTER CONVERSION OPTIONS

The remaining options determine whether a letter conversion will take place. The general syntax is a conversion option is enabled if beginning with '+', and disabled if beginning with '-'.

- +all Enable all conversions. This is the default.
- all Disable all conversions. Individual conversions can be selected after the default of all conversions is disabled.
- +ae, +ash Convert "ae" to the ash (combined a+e) character, in upper- or lower-case as appropriate: Unicode code points U+00C6 and U+00E6, respectively.
- ae, -ash Do not convert "ae" to the ash character.
- +oe, +ethel Convert "oe" to the ethel (combined o+e) character, in upper- or lower-case as appropriate: Unicode code points U+0152 and U+0153, respectively.
- oe, -ethel Do not convert "oe" to the ethel character.
- +g Convert "g" to upper- or lower-case insular "g": Unicode code points U+A77D and U+1D79, respectively.
- g Do not convert "g" to insular "g".
- +gh, +yogh Convert "gh" to the upper- or lower-case yogh character: Unicode code points U+021C and U+021D, respectively.
- gh, -yogh Do not convert "gh" to the yogh character.
- +insular Convert to insular forms of upper- and lower-case D, F, G, R, S, and T.
- insular Do not convert to insular forms of upper- and lower-case D, F, G, R, S, and T. Note that insular "G" and "g" will still be converted if the "+g" option is active.
- +s Convert initial and medial lower-case 's' to the long s form, Unicode code point U+017F.
- s Do not convert initial or medial lower-case 's' to the long s form.
- +thorn Convert unvoiced "th" to upper- or lower-case thorn, Unicode code points U+00DE or U+00FE, resp. Also, if the "+eth" option is not chosen to convert voiced "th" to eth, then voiced "th" will also be converted to thorn with this option selected.
- thorn Do not convert "th" to thorn.

+eth	Convert voiced "th" to upper- or lower-case eth, Unicode code points U+00D0 or U+00F0, resp.
-eth	Do not convert voiced "th" to eth.
+th	Convert voiced and unvoiced "th" to eth and thorn, respectively. See the options "+thorn" and "+eth" for details.
-th	Do not convert voiced or unvoiced "th" to eth or thorn.
+w	Convert 'w' to the ancient letter upper- or lower-case wynn, Unicode code points U+01F7 or U+01BF, resp.

ESCAPE SEQUENCES

olden will always convert the following escape sequences within a file before performing any other conversion. These sequences cannot be overridden by command line arguments. They are modeled after those available in TeX and LaTeX. Accents show the letter "a" inside each "{}" pair as an example letter to be accented; Unicode code points and Unicode names appear for each sequence:

\`{a}	U+0300, COMBINING GRAVE ACCENT.
\' {a}	U+0301, COMBINING ACUTE ACCENT.
\~{a}	U+0302, COMBINING CIRCUMFLEX ACCENT.
\~{a}	U+0303, COMBINING TILDE.
\={a}	U+0304, COMBINING MACRON.
\u{a}	U+0306, COMBINING BREVE.
\. {a}	U+0307, COMBINING DOT ABOVE.
\" {a}	U+0308, COMBINING DIAERESIS.
\r{a}	U+030A, COMBINING RING ABOVE.
\H{a}	U+030B, COMBINING DOUBLE ACUTE ACCENT.
\v{a}	U+030C, COMBINING CARON.
\U{a}	U+030E, COMBINING DOUBLE VERTICAL LINE ABOVE.
\G{a}	U+030F, COMBINING DOUBLE GRAVE ACCENT.
\d{a}	U+0323, COMBINING DOT BELOW.
\c{a},	U+0327, COMBINING CEDILLA. Note that the "C" and "c" with a cedilla generate U+00C7, LATIN CAPITAL LETTER C WITH CEDILLA, and U+00E7, LATIN SMALL LETTER C WITH CEDILLA, respectively.
\k{a}	U+0328, COMBINING OGONEK. Note that the Saxon letter "E" and "e" with a cedilla generate U+0118, LATIN CAPITAL LETTER E WITH OGONEK, and U+0119, LATIN SMALL LETTER E WITH OGONEK, respectively.
\b{a}	U+0331, COMBINING MACRON BELOW.
\t{ae}	U+0361, COMBINING DOUBLE INVERTED BREVE; also referred to as a "tie".
\AA{ }	U+00C5, LATIN CAPITAL LETTER A WITH RING ABOVE.
\aa{ }	U+00E5, LATIN SMALL LETTER A WITH RING ABOVE.
\AE{ }	U+00C6, LATIN CAPITAL LETTER AE.
\ae{ }	U+00E6, LATIN SMALL LETTER AE.
\DH{ }	U+00D0, LATIN CAPITAL LETTER ETH.
\dh{ }	U+00F0, LATIN SMALL LETTER ETH.

<code>\DJ{}</code>	U+0110, LATIN CAPITAL LETTER D WITH STROKE.
<code>\dj{}</code>	U+0111, LATIN SMALL LETTER D WITH STROKE.
<code>\I{}</code>	ASCII "I".
<code>\i{}</code>	U+0131, LATIN SMALL LETTER DOTLESS I. Note that for any accent that appears above "i", olden automatically converts "i" to the dotless form.
<code>\J{}</code>	ASCII "J".
<code>\j{}</code>	U+0237, LATIN SMALL LETTER DOTLESS J. Note that for any accent that appears above "j", olden automatically converts "j" to the dotless form.
<code>\L{}</code>	U+0141, LATIN CAPITAL LETTER L WITH STROKE.
<code>\l{}</code>	U+0142, LATIN SMALL LETTER L WITH STROKE.
<code>\NG{}</code>	U+014A, LATIN CAPITAL LETTER ENG.
<code>\ng{}</code>	U+014B, LATIN SMALL LETTER ENG.
<code>\O{}</code>	U+00D8, LATIN CAPITAL LETTER O WITH STROKE.
<code>\o{}</code>	U+00F8, LATIN SMALL LETTER O WITH STROKE.
<code>\OE{}</code>	U+0152, LATIN CAPITAL LIGATURE OE.
<code>\oe{}</code>	U+0153, LATIN SMALL LIGATURE OE.
<code>\SS{}</code>	ASCII "SS".
<code>\ss{}</code>	U+00DF, LATIN SMALL LETTER SHARP S.
<code>\TH{}</code>	U+00DE, LATIN CAPITAL LETTER THORN.
<code>\th{}</code>	U+00FE, LATIN SMALL LETTER THORN.
<code>\Thaet{}</code>	U+A764, LATIN CAPITAL LETTER THORN WITH STROKE.
<code>\thaet{}</code>	U+A765, LATIN SMALL LETTER THORN WITH STROKE.
<code>\textexclamdown{}</code>	U+00A1, INVERTED EXCLAMATION MARK.
<code>\textquestiondown{}</code>	U+00BF, INVERTED QUESTION MARK.
<code>\textordfeminine{}</code>	U+00AA, FEMININE ORDINAL INDICATOR.
<code>\textordmasculine{}</code>	U+00BA, MASCULINE ORDINAL INDICATOR.

EXAMPLE

Sample usage:

```
olden -all +s +th -i my_input_file.txt -o my_output_file.utf8
```

EXIT STATUS

olden will exit with a status of `EXIT_SUCCESS` if the command line options are all valid. The exit status will be `EXIT_FAILURE` if an invalid option was requested or there is an error opening files specified for input or output.

FILES

ASCII text input files; UTF-8 output files.

AUTHOR

olden was written by Paul Hardy.

LICENSE

olden is Copyright © 2019 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.