

NAME

set_unicharset_properties – set properties about the unichars

SYNOPSIS

set_unicharset_properties --U *input_unicharsetfile* --script_dir */path/to/langdata* --O *output_unicharsetfile*

DESCRIPTION

set_unicharset_properties(1) reads a unicharset file, puts the result in a UNICHARSET object, fills it with properties about the unichars it contains and writes the result back to another unicharset file.

OPTIONS

--script_dir */path/to/langdata*

(Input) Specify the location of directory for universal script unicharsets and font xheights (type:string default:)

--U *unicharsetfile*

(Input) Specify the location of the unicharset to load as input.

--O *unicharsetfile*

(Output) Specify the location of the unicharset to be written with updated properties.

HISTORY

set_unicharset_properties(1) was first made available for tesseract version 3.03.

RESOURCES

Main web site: <https://github.com/tesseract-ocr> Information on training:
<https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract>

SEE ALSO

tesseract(1)

COPYING

Copyright (C) 2012 Google, Inc. Licensed under the Apache License, Version 2.0

AUTHOR

The Tesseract OCR engine was written by Ray Smith and his research groups at Hewlett Packard (1985–1995) and Google (2006–present).