

NAME

merge_unicharsets – Simple tool to merge two or more unicharsets.

SYNOPSIS

merge_unicharsets *unicharset-in-1 ... unicharset-in-n unicharset-out*

DESCRIPTION

merge_unicharsets(1) is a simple tool to merge two or more unicharsets. It could be used to create a combined unicharset for a script-level engine, like the new Latin or Devanagari.

IN/OUT ARGUMENTS

unicharset-in-1

(Input) The name of the first unicharset file to be merged.

unicharset-in-n

(Input) The name of the nth unicharset file to be merged.

unicharset-out

(Output) The name of the merged unicharset file.

HISTORY

merge_unicharsets(1) was first made available for tesseract4.00.00alpha.

RESOURCES

Main web site: <https://github.com/tesseract-ocr> Information on training tesseract LSTM:
<https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract-4.00>

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).