

**NAME**

classifier\_tester – for \*legacy tesseract\* engine.

**SYNOPSIS**

**classifier\_tester** -U *unicharset\_file* -F *font\_properties\_file* -X *xheights\_file* -classifier *x* -lang *lang*  
 [-output\_trainer *trainer*] \*.tr

**DESCRIPTION**

classifier\_tester(1) runs Tesseract in a special mode. It takes a list of .tr files and tests a character classifier on data as formatted for training, but it doesn't have to be the same as the training data.

**IN/OUT ARGUMENTS**

a list of .tr files

**OPTIONS**

-l *lang*

(Input) three character language code; default value *eng*.

-classifier *x*

(Input) One of "pruner", "full".

-U *unicharset*

(Input) The unicharset for the language.

-F *font\_properties\_file*

(Input) font properties file, each line is of the following form, where each field other than the font name is 0 or 1:

\*font\_name\* \*italic\* \*bold\* \*fixed\_pitch\* \*serif\* \*fraktur\*

-X *xheights\_file*

(Input) x heights file, each line is of the following form, where xheight is calculated as the pixel x height of a character drawn at 32pt on 300 dpi. [ That is, if base x height + ascenders + descenders = 133, how much is x height? ]

\*font\_name\* \*xheight\*

-output\_trainer *trainer*

(Output, Optional) Filename for output trainer.

**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).