

## NAME

`ambiguous_words` – generate sets of words Tesseract is likely to find ambiguous

## SYNOPSIS

**`ambiguous_words`** [-l lang] *TESSDATADIR WORDLIST AMBIGUOUSFILE*

## DESCRIPTION

`ambiguous_words(1)` runs Tesseract in a special mode, and for each word in word list, produces a set of words which Tesseract thinks might be ambiguous with it. *TESSDATADIR* must be set to the absolute path of a directory containing *tessdata/lang.traineddata*.

## 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).