Tasks - [Conan](https://image6.thenewslens.com/2021/5/fsgtowauouizwxgrpcncyri6n511at.jpg?auto=compress&fit=crop&h=648&q=85&updated_at=2021-05-03-12-31-09&w=1080) Team

# Question 1

\*\* We wrote > instead of indents. > = one indent, >> = two indents,…

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Test Case ID** | **Feature Breakdown** | **Probability [P][0-1]** | **Loss Factor [L] [1-10]** | **Exposure [E]** | **Comment** | **Priority level** | **Notes/Question** |
|  | **ConfigFile** |  |  |  |  |  |  |
|  | > Passing wrong file path |  |  |  | Make sure it uses default values |  |  |
|  | >Passing invalid file (not in english/gibberish) |  |  |  |  |  |  |
|  | > Passing valid file |  |  |  |  |  |  |
|  | >>Passing empty file |  |  |  |  |  |  |
|  | >>Passing a file with one configuration |  |  |  |  |  |  |
|  | >> Configurations without values |  |  |  | Does it ignore them, or error message? |  |  |
|  | >>Configurations with multiple values |  |  |  |  |  |  |
|  | >> Make sure each configuration parameter is detectable |  |  |  | Is there a need to test all the configurations? |  |  |
|  | >> Configuration Parameters |  |  |  |  |  |  |
|  | >>> Writing configuration parameters more than once |  |  |  | Which value is chosen? |  |  |
|  | **-c flag** |  |  |  |  |  |  |
|  | > Pass an undefined configuration parameter |  |  |  |  |  |  |
|  | > Pass the same configuration parameter more than once |  |  |  | Which value is chosen |  |  |
|  | >Pass an invalid parameter for a defined configuration |  |  |  |  |  |  |
|  | >Pass multiple configurations with multiple parameters |  |  |  |  |  |  |
|  | Command line+ configuration File |  |  |  |  |  |  |
|  | >>Pass **different** parameters/configurations on both the command line and the configuration file |  |  |  |  |  | **Should give different outputs** |
|  | >>Pass **identical** parameters/configurations on both the command line and the configuration file |  |  |  |  |  | **Should give identical outputs** |

# Question 2 - Page Separator

## What it does

Page separator is used to define a string or a character that will work as a separator for the OCR result after every image\page.

The page separator is written on a new line after the result of the OCR of an image, and after it there would be the result of the next image in the same line.

* Using the empty string as a page separator simply omits separators.
* The maximum length of a separator is 254 characters.

## Bugs

### New line character

The page separator configuration parameter doesn’t support the new line character. When passing the new line character “\n” to the page separator parameter, Tesseract doesn’t interpret it as a new line instead it writes “\n” to the output.

Steps to reproduce

* Perform OCR on a number of images at the same time.
* Use “\n” as the page separator parameter.
* Check the output.

Expected result: The output file contains the texts of each image separated by a new line.

Actual result: Result will contain the “\n” and right after that, the text of the second image.

### No space between page separator and the output

The results of the OCR continue right after the page separator without a space between the page separator and the result. This could create confusion in the actual results, where the results will be mixed with the page separator.

Steps to reproduce

* Perform OCR on a number of images at the same time.
* Use any string other than the empty string as a page separator.
* Check the output.

Expected result: The actual result will be separated from the page separator string using a space or a new line.

Actual result: The result of the images is connected to the page separator without any separators.

# Question 3 - Metamorphic Relations

1. Given image X and OCR output Y:

TEST CASE: blacklisting an appearing letter σ within a given output text, outputs at most the same size as the original text.

If we take one letter that was present in the output and add it to the blacklist, we expect the new result to have less text, or to be specific, to not include this letter.

|  |  |  |
| --- | --- | --- |
| Relation | Metamorphic test | Expected result |
| Output Y has a letter X | Blacklist X |  |
| Output Y had a blocked letter X | Unblack list X |  |
| Output Y has a letter X | Whitelist X |  |
| Output Y has multiple letters | Whitelist some combination |  |

These tests can be generated for every X and every letter or combination of letters, so we got a wide range of tests.

In other words:

* Given X, Y: if we whitelist the output Y or a chunk of it, we should have a text that is equal to Y or to the chunk.
* If we take one letter that was present in the output and add it to the blacklist, we expect the new result to have less text, or to be specific, to not include this letter.

1. Given image X, and OCR output Y:

TEST CASE: Rotating the image by must give the same output.

|  |  |  |
| --- | --- | --- |
| Relation | Metamorphic test | Expected result |
| Given X, Y | Rotate X by | and the text is identical |
| Given X, Y | Rotate X by | and the text is identical |
| Given X, Y | Rotate X by | and the text is identical |

Rotate X 90,180,270 angle and compare the images, the rotated image should have the same result (Y)

1. Given image X and OCR output Y:

TEST CASE: blacklisting an appearing letter σ within a given output text, outputs at most the same number of different characters as the original text.

|  |  |  |
| --- | --- | --- |
| Relation | Metamorphic test | Expected result |
| Output Y has multiple letters | Whitelist some set of characters |  |
| Output Y has multiple letters | blacklist some set of characters |  |

These tests can be generated for every X and every letter or combination of letters, so we got a wide range of tests.

In other words:

* Given Output Y: if we white/black-list any set of characters we should get Y’ in which all characters appear also in the original Y

# Question 4

# Q 4.1

|  |  |
| --- | --- |
| Length of text | equivalence classes :   1. Empty 2. Not empty, <One line 3. 1< number of lines |
| Position of text in the image | We can divided the image to nine squares (1-9, like a tic-tac-toe board) and have the following equivalence classes, based on **in which square the text begins**:   |  |  |  | | --- | --- | --- | | Square 1 | Square 2 | Square 3 | | Square 4 | Square 5 | Square 6 | | Square 7 | Square 8 | Square 9 |   0. Empty   1. Square 1 2. Square 2 3. … 4. Square 9 |
| Language | In this exercise we deal with four language so we should consider every language a separate equivalence class:   1. English 2. Hebrew 3. Swedish 4. Spanish |
| Space between words and lines | Equivalence classes - spaces between words:   1. No spaces between words 2. One space between words only 3. 1<Number of space between some words   Equivalence classes - spaces between lines:   1. No spaces between lines. 2. One space between lines 3. 1 < Number of spaces between lines |
| Text direction (angle) | Equivalence classes :   1. 0 degrees 2. 90 degrees 3. 180 degrees 4. 270 degrees |
| The content of the text | Equivalence classes:   1. Empty 2. Non-empty |

# Q 4.2 - Test strategy

The test strategy for Tesseract -an OCR application-:

First, to make it crystal clear, we will not attempt to cover every single test there is, however, our goal is to obtain the maximum coverage possible. And to do so, we will build various test cases, partitioned for specific goals (such as validating correct behaviour for basic tests, generic font sizes, language recognition, etc...).

Given the assumptions list about what tesseract can and cannot do, we came up with the following test strategy:

Our testing cycle contains tests for each of the following case:

* The empty image
* Basic tests
* Fonts
* Font sizes
* The configuration parameters
* Variations in the text in the image:
  + Page separators
  + Length of text
  + Position of text in the image
  + Language
  + Space between words and lines
  + The content of the text
* Combinations of the above:
  + Font size combinations
  + Font combinations
  + Language combinations

The logic behind the strategy comes from the maximum likelihood of events, we focused on test cases that we think have the highest occurrence rate to the average user, to make sure that on average the application works, this can be seen in the basic tests, where we try one language, with one or two lines, that are not rotated and clear -not close to the edges and with no overlapping lines\words-.

Afterwards, we started with covering the fonts that are supported, the logic here is to take a font size that is big enough, with the same sentence written in the image and the same spacing, so that the only factor to the test is the different font styles that we have, we chose the font size 38, with all the supported fonts. As we figured that failing any of these tests is bad for the application as it assumes 100% recognition rate for these fonts.

A similar approach was used to test the font sizes, where we wrote the same sentence using the same font style with no overlapping\rotation to check the font sizes from 8 to 72.

for the language recognition tests, we conducted simple tests; for each language, we have two images with the alphabet, one in uppercase and one in lower, using the same font, the idea behind this is that, if the tesseract can recognize the alphabet correctly, then he can probably recognize the language well. For languages like Hebrew, we also throwed a sentence test into the mix, the logic behind it is that letters are written differently according to where they are in the word, (last letter etc.), also, for Spanish, we had a test for the upside question mark that is present in the language, and an accent test, using language diacritics, finally, we used one numbers image and one symbol image just to cover as much as possible.

Moreover, we conducted tests to different text positioning and rotations using the same font/sentence/language, to check the invariability of rotations promised in the requirements sheet.

Lastly, to get as much coverage as possible, we conducted a combination test, where for one image we could run various configurations with different parameters and languages, this way using at most 300 images, we could build thousands of tests.

It’s worth noting that our tests almost don’t overlap, since we think it’s a waste of computational power and time, as too much overlapping does not help with our goal (maximum coverage) specially when we have a performance limit of 300 images and 15 minutes of run time. adopting this technique (all of the above) in our implementation will result in the maximum coverage possible, as constraining a set of parameters while checking others allows us to generate way over 300 test combinations, that's because 300 tests (i.e., 1 test per image) could never cover all the combinations of the parameters. As a result, our orientation towards bugs is “all the bugs that we can catch in the combinations that we chose to test” and according to human logic, the more common the combination under test the more dangerous the bug and more cumbersome, and the less probable the combination the rarer the bug, and the less dangerous it is -in most cases-, and we are back to the maximum likelihood strategy that we mentioned above.

## 

# Q 4.3 - Bug report

## Most critical bugs

1. Tesseract fails to recognise texts written in font size 12 or less (8 - 12). Font size 12 and 11 are one of the most popular font sizes that are widely used. Not being able to recognize them is a critical bug in Tesseract.   
     
   Steps to reproduce:
   1. Run OCR on images\font\_size\_images\font\_size\_12\_test.png.
   2. Compare results with image contents.
   3. You’ll see that the results are different and the contents are not recognized as needed.

Expected: The OCR will detect the contents of the image.

Actual: The OCR returns an empty file.

1. Tesseract fails to recognize the “David” Hebrew font correctly, which is one of the most popular fonts used for the Hebrew language. Some of the Hebrew letters are recognized by Tesseract as other letters when using the David font.   
     
   Steps to reproduce:
   1. Run OCR on images\fonts\_images\david\_38\_hebrew.png.
   2. Compare results with image contents.
   3. You’ll see that the results are different and the contents are not recognized as needed.

Expected: The OCR will detect the contents of the image.

Actual: The OCR doesn't detect all the Hebrew alphabets correctly.

1. Tesseract fails to recognize text in images where the text is in English, Spanish and Swedish at the same time. It fails to distinguish between the different languages.  
     
   Steps to reproduce:
   1. Run OCR on images\language\_combinations\english\_spanish\_swedish with the parameters “-l spa+swe+eng”
   2. Compare the results with the image contents.
   3. You’ll see that the results are different and the contents are not recognized as needed.

Expected: Each language will be detected correctly.

Actual: The output file is empty.

1. Using the unblacklist configuration parameter affects the whitelist configuration parameter. It causes Tesseract to detect more letters compared to what is found in the whitelist.

Steps to reproduce:

* 1. Run OCR on images\config\_images\english\_28\_letters\_double with the parameters “-c tessedit\_char\_whitelist="mosa" -c tessedit\_char\_unblacklist="karemos"”
  2. Compare the results with the image contents.
  3. You’ll see that the results are different and the contents are not recognized as needed.

Expected: The output will only contain letters from the whitelist.

Actual: The output contains letters from inside and outside the whitelist.

1. Tesseract doesn’t detect the number of line spaces between two lines of text.   
     
   Steps to reproduce:
   1. Run OCR on images\line\_word\_spacing\8\_line\_spaces.png.
   2. Compare the results with the image contents.
   3. You’ll see that the number of line spaces in the result doesn’t don’t match the content of the image.

Expected: The same number of line spaces as the image.

Actual: Doesn’t pay attention to space lines.

### All the bugs

|  |  |  |
| --- | --- | --- |
| **Description** | **Test #** | **Command** |
| Tesseract failed to identify font size 8 | 3 | tesseract "images\font\_size\_images\font\_size\_8\_test.png" "images\font\_size\_images\font\_size\_8\_test" -l eng -c page\_separator="" |
| Tesseract failed to identify font size 9 | 4 | tesseract "images\font\_size\_images\font\_size\_9\_test.png" "images\font\_size\_images\font\_size\_9\_test" -l eng -c page\_separator="" |
| Tesseract failed to identify font size 10 | 5 | tesseract "images\font\_size\_images\font\_size\_10\_test.png" "images\font\_size\_images\font\_size\_10\_test" -l eng -c page\_separator="" |
| Tesseract failed to identify font size 11 | 6 | tesseract "images\font\_size\_images\font\_size\_11\_test.png" "images\font\_size\_images\font\_size\_11\_test" -l eng -c page\_separator="" |
| Tesseract failed to identify font size 12 | 7 | tesseract "images\font\_size\_images\font\_size\_12\_test.png" "images\font\_size\_images\font\_size\_12\_test" -l eng -c page\_separator="" |
| Tesseract failed to identify font impact | 74 | tesseract "images\fonts\_images\impact\_38\_english.png" "images\fonts\_images\impact\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font Inkfree | 77 | tesseract "images\fonts\_images\Inkfree\_38\_english.png" "images\fonts\_images\Inkfree\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font lvnm | 79 | tesseract "images\fonts\_images\lvnm\_38\_english.png" "images\fonts\_images\lvnm\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font lvnmbd | 80 | tesseract "images\fonts\_images\lvnmbd\_38\_english.png" "images\fonts\_images\lvnmbd\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font lucon | 83 | tesseract "images\fonts\_images\lucon\_38\_english.png" "images\fonts\_images\lucon\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font corbel | 90 | tesseract "images\fonts\_images\corbel\_38\_english.png" "images\fonts\_images\corbel\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font courier | 92 | tesseract "images\fonts\_images\cour\_38\_english.png" "images\fonts\_images\cour\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font mrimac | 93 | tesseract "images\fonts\_images\mriamc\_38\_english.png" "images\fonts\_images\mriamc\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font ms gothic | 95 | tesseract "images\fonts\_images\msgothic\_38\_english.png" "images\fonts\_images\msgothic\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font ebrima | 98 | tesseract "images\fonts\_images\ebrima\_38\_english.png" "images\fonts\_images\ebrima\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font rod | 99 | tesseract "images\fonts\_images\rod\_38\_english.png" "images\fonts\_images\rod\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font forte | 100 | tesseract "images\fonts\_images\FORTE\_38\_english.png" "images\fonts\_images\FORTE\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font framd | 102 | tesseract "images\fonts\_images\framd\_38\_english.png" "images\fonts\_images\framd\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font segoesc | 103 | tesseract "images\fonts\_images\segoesc\_38\_english.png" "images\fonts\_images\segoesc\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font gabriola | 106 | tesseract "images\fonts\_images\Gabriola\_38\_english.png" "images\fonts\_images\Gabriola\_38\_english" -l eng -c page\_separator="" |
| Tesseract failed to identify font ahronbd in hebrew | 110 | tesseract "images\fonts\_images\ahronbd\_38\_hebrew.png" "images\fonts\_images\ahronbd\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify gishabd in hebrew | 113 | tesseract "images\fonts\_images\gishabd\_38\_hebrew.png" "images\fonts\_images\gishabd\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify calibri in hebrew | 114 | tesseract "images\fonts\_images\calibri\_38\_hebrew.png" "images\fonts\_images\calibri\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify courier in hebrew | 121 | tesseract "images\fonts\_images\cour\_38\_hebrew.png" "images\fonts\_images\cour\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify mriamc in hebrew | 122 | tesseract "images\fonts\_images\mriamc\_38\_hebrew.png" "images\fonts\_images\mriamc\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify david in hebrew | 123 | tesseract "images\fonts\_images\david\_38\_hebrew.png" "images\fonts\_images\david\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify segoeui in hebrew | 128 | tesseract "images\fonts\_images\segoeui\_38\_hebrew.png" "images\fonts\_images\segoeui\_38\_hebrew" -l heb -c page\_separator="" |
| Tesseract failed to identify text rotated in 90 degrees | 146 | tesseract "images\position\_images\rot\_90.png" "images\position\_images\rot\_90" -l eng -c page\_separator="" |
| Tesseract failed to identify text rotated in 180 degrees | 147 | tesseract "images\position\_images\rot\_180.png" "images\position\_images\rot\_180" -l eng -c page\_separator="" |
| Tesseract failed to identify text rotated in 270 degrees | 148 | tesseract "images\position\_images\rot\_270.png" "images\position\_images\rot\_270" -l eng -c page\_separator="" |
| Tesseract failed to identify symbols | 150 | tesseract "images\language\_images\symbols\_image.png" "images\language\_images\symbols\_image" -l eng -c page\_separator="" |
| Tesseract failed to identify lower case spanish alphabets | 153 | tesseract "images\language\_images\spanish\_alphabets\_lower.png" "images\language\_images\spanish\_alphabets\_lower" -l spa -c page\_separator="" |
| Tesseract failed to identify upper case spanish alphabets | 154 | tesseract "images\language\_images\spanish\_alphabets\_upper.png" "images\language\_images\spanish\_alphabets\_upper" -l spa -c page\_separator="" |
| Tesseract failed to identify spanish letters with accents | 155 | tesseract "images\language\_images\spanish\_accent.png" "images\language\_images\spanish\_accent" -l spa -c page\_separator="" |
| Tesseract failed to identify lower case swedish alphabets | 157 | tesseract "images\language\_images\swedish\_alphabets\_lower.png" "images\language\_images\swedish\_alphabets\_lower" -l swe -c page\_separator="" |
| Tesseract failed to identify upper case swedish alphabets | 158 | tesseract "images\language\_images\swedish\_alphabets\_upper.png" "images\language\_images\swedish\_alphabets\_upper" -l swe -c page\_separator="" |
| Tesseract failed to identify hebrew sentence | 162 | tesseract "images\language\_images\hebrew\_sentence.png" "images\language\_images\hebrew\_sentence" -l heb -c page\_separator="" |
| Tesseract failed to identify image with 4 line spaces | 166 | tesseract "images\line\_word\_spacing\4\_line\_spaces.png" "images\line\_word\_spacing\4\_line\_spaces" -l eng -c page\_separator="" |
| Tesseract failed to identify image with 8 line spaces | 167 | tesseract "images\line\_word\_spacing\8\_line\_spaces.png" "images\line\_word\_spacing\8\_line\_spaces" -l eng -c page\_separator="" |
| Tesseract failed to identify an image full of text | 169 | tesseract "images\text\_length\_images\12\_lines\_image.png" "images\text\_length\_images\12\_lines\_image" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 170 | tesseract "images\font\_size\_combinations\sizes\_14\_11.png" "images\font\_size\_combinations\sizes\_14\_11" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 171 | tesseract "images\font\_size\_combinations\sizes\_16\_11.png" "images\font\_size\_combinations\sizes\_16\_11" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 172 | tesseract "images\font\_size\_combinations\sizes\_18\_11.png" "images\font\_size\_combinations\sizes\_18\_11" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 173 | tesseract "images\font\_size\_combinations\sizes\_20\_11.png" "images\font\_size\_combinations\sizes\_20\_11" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 174 | tesseract "images\font\_size\_combinations\sizes\_11\_14.png" "images\font\_size\_combinations\sizes\_11\_14" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 175 | tesseract "images\font\_size\_combinations\sizes\_11\_16.png" "images\font\_size\_combinations\sizes\_11\_16" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 176 | tesseract "images\font\_size\_combinations\sizes\_11\_18.png" "images\font\_size\_combinations\sizes\_11\_18" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 177 | tesseract "images\font\_size\_combinations\sizes\_11\_20.png" "images\font\_size\_combinations\sizes\_11\_20" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with texts in different sizes | 178 | tesseract "images\font\_size\_combinations\small\_big\_sizes.png" "images\font\_size\_combinations\small\_big\_sizes" -l eng -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 184 | tesseract "images\language\_combinations\english\_and\_spanish.png" "images\language\_combinations\english\_and\_spanish" -l eng+spa -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 185 | tesseract "images\language\_combinations\english\_and\_spanish.png" "images\language\_combinations\english\_and\_spanish" -l spa+eng -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 186 | tesseract "images\language\_combinations\english\_and\_swedish.png" "images\language\_combinations\english\_and\_swedish" -l eng+swe -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 188 | tesseract "images\language\_combinations\spanish\_and\_swedish.png" "images\language\_combinations\spanish\_and\_swedish" -l spa+swe -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 189 | tesseract "images\language\_combinations\spanish\_and\_swedish.png" "images\language\_combinations\spanish\_and\_swedish" -l swe+spa -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 190 | tesseract "images\language\_combinations\english\_spanish\_swedish" "images\language\_combinations\english\_spanish\_swedish" -l eng+spa+swe -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 191 | tesseract "images\language\_combinations\english\_spanish\_swedish" "images\language\_combinations\english\_spanish\_swedish" -l eng+swe+spa -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 192 | tesseract "images\language\_combinations\english\_spanish\_swedish" "images\language\_combinations\english\_spanish\_swedish" -l spa+eng+swe -c page\_separator="" |
| Tesseract failed to identify an image with different languages | 193 | tesseract "images\language\_combinations\english\_spanish\_swedish" "images\language\_combinations\english\_spanish\_swedish" -l spa+swe+eng -c page\_separator="" |
| Tesseract failed to identify text when using configuration parameters | 198 | tesseract "images\fonts\_images\ariblk\_38\_english.png" "images\fonts\_images\ariblk\_38\_english" -l eng -c page\_separator="" -c tessedit\_char\_blacklist="abcdefghijklnopqrstuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 199 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_blacklist="karem" |
| Tesseract failed to identify text when using configuration parameters | 201 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="m" |
| Tesseract failed to identify text when using configuration parameters | 202 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="mosa" |
| Tesseract failed to identify text when using configuration parameters | 203 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="abcdefghijklmnopqrstuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 204 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="m" -c tessedit\_char\_blacklist="k" |
| Tesseract failed to identify text when using configuration parameters | 205 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="m" -c tessedit\_char\_blacklist="m" |
| Tesseract failed to identify text when using configuration parameters | 206 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="m" -c tessedit\_char\_blacklist="mosa" |
| Tesseract failed to identify text when using configuration parameters | 207 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="m" -c tessedit\_char\_blacklist="abcdefghijklmnopqrstuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 208 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="karem" -c tessedit\_char\_blacklist="abcdefghijklmnopqrstuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 209 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_unblacklist="abcdefghijklmnopqrstuvwxyz" -c tessedit\_char\_blacklist="karemos" |
| Tesseract failed to identify text when using configuration parameters | 210 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="karem" |
| Tesseract failed to identify text when using configuration parameters | 211 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="abcdefghijklmnopqrstuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 212 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="abcdefghijklmnopqrstuvwxyz" -c tessedit\_char\_blacklist="mosa" |
| Tesseract failed to identify text when using configuration parameters | 213 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="mosa" -c tessedit\_char\_blacklist="bcdefghijklnpqrtuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 214 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="karem" -c tessedit\_char\_blacklist="karem" |
| Tesseract failed to identify text when using configuration parameters | 215 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="mosa" -c tessedit\_char\_blacklist="bcdefghijklnpqrtuvwxyz" |
| Tesseract failed to identify text when using configuration parameters | 216 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="mosa" -c tessedit\_char\_unblacklist="kfc" |
| Tesseract failed to identify text when using configuration parameters | 217 | tesseract "images\config\_images\english\_28\_letters\_double.png" "images\config\_images\english\_28\_letters\_double" -l eng -c page\_separator="" -c tessedit\_char\_whitelist="mosa" -c tessedit\_char\_unblacklist="karemos" |
| Tesseract failed to use the page separator correctly | 218 | tesseract "images\page\_seprator\_tools\basic\_test.txt" "images\page\_seprator\_tools\basic\_test.txt" -l eng -c page\_separator="aaaaaaaaaa" |
| Tesseract failed to use the page separator correctly | 219 | tesseract "images\page\_seprator\_tools\complex\_test.txt" "images\page\_seprator\_tools\complex\_test.txt" -l eng -c page\_separator="aaaaaaaaaa" |
| Tesseract failed to use the page separator correctly | 220 | tesseract "images\page\_seprator\_tools\complex\_test.txt" "images\page\_seprator\_tools\complex\_test.txt" -l eng -c page\_separator="aaaaaaaaaa" -c tessedit\_char\_blacklist="a" |
| Tesseract failed to use the page separator correctly | 221 | tesseract "images\page\_seprator\_tools\complex\_test.txt" "images\page\_seprator\_tools\complex\_test.txt" -l eng -c page\_separator="awawwaaa waaa awwwa wwa ww aawwwaa awwwaw awwwawawawawaaawawaw awaaawaw awawaawaawawawawawawaawawaawaawawawawaaawawaawawawawawwaawawwaawawawwaawawawaw" |
| Tesseract failed to use the page separator correctly | 222 | tesseract "images\page\_seprator\_tools\complex\_test.txt" "images\page\_seprator\_tools\complex\_test.txt" -l eng -c page\_separator="aaaaaaaaaa" -c tessedit\_char\_whitelist="thisfulineo" |