Captcha Analyzer

# Picture to Text

## Steps

1. Using Selenium, screenshot the captcha element
2. Remove the frame & Filter the noise from the captcha with the selected method
3. Send filtered picture to the API model ("Text Captcha Breaker")
4. Correct the API results as to the specific captcha
5. Send the result back to the site and check if passed
6. Save the captcha, the filtered captcha and the filtering process with the pass/fail result in history folder for future ML improvements
7. Redo the above steps if the site is asking for another captcha

(Two captchas in case of using zone-h search filter and more times in case of failing)

1. Send the ZHE & PHPSESSID to the calling program to able it to run without the captcha requirement

## Tested Methods

1. Median
2. Median -> Mask with original
3. Median -> Dilation -> Erosion
4. Median -> Mask with original -> Dilation -> Erosion
5. Dilation -> Erosion
6. Median -> Mask with original -> Median -> Mask with original -> Enlarge by 300% -> Dilation
7. Median -> Mask with original -> Median -> Mask with original -> Enlarge by 300% -> Dilation -> Erosion
8. Noise color filtering
9. Text color filtering

\* Each method also tested with and without API result correction

## API Results Correction (optional)

1. Small letters detection: Convert to similar capitals

|  |  |
| --- | --- |
| **Chars Conversion Table** | |
| e | C |
| l | I |

\* "C" example from a captcha:



1. Numbers detection: Convert to similar letters

|  |  |
| --- | --- |
| **Numbers Conversion Table** | |
| 0 | O |
| 1 | I |
| 2 | Z |
| 3 | B |
| 4 | A |
| 5 | S |
| 6 | E |
| 7 | X |
| 8 | B |
| 9 | O |