

Resolver el CAPTCHA de Amazon

El reconocimiento óptico de caracteres (ROC), generalmente conocido como reconocimiento de caracteres y expresado con frecuencia con la sigla OCR (del inglés OpticalCharacter Recognition), es un proceso dirigido a la digitalización de textos, los cuales identifican automáticamente a partir de una imagen símbolos o caracteres que pertenecen a un determinado alfabeto, para luego almacenarlos en forma de datos. Así podremos interactuar con estos mediante un programa de edición de texto o similar.

```
#Descargar la imagen del CAPTCHA de Amazon
$url="https://www.amazon.es/errors/validateCaptcha"
$result = Invoke-WebRequest $url
$fichero = (Get-date).Ticks
Start-BitsTransfer ($result.Images.src | Select-String
"captcha") -Destination ([String]$fichero+".jpg")
Start-Process ([String]$fichero+".jpg")
```

#Convertir la imagen CAPTCHA a texto (el resultado obtenido no es positivo en la mayoría de las ocasiones)

#Tesseract OCR

#Tesseract is probably the most accurate open source OCR engine available. Combined with the Leptonica Image Processing Library it can read a wide variety of image formats and convert them to text in over 60 languages. It was one of the top 3 engines in the 1995 UNLV Accuracy test. Between 1995 and 2006 it had little work done on it, but since then it has been improved extensively by Google. It is released under the Apache License 2.0.

```
. "C:\Program Files (x86)\Tesseract-OCR\tesseract.exe"
([String]$fichero+".jpg") ([String]$fichero)
gc ([String]$fichero+".txt")
```

