

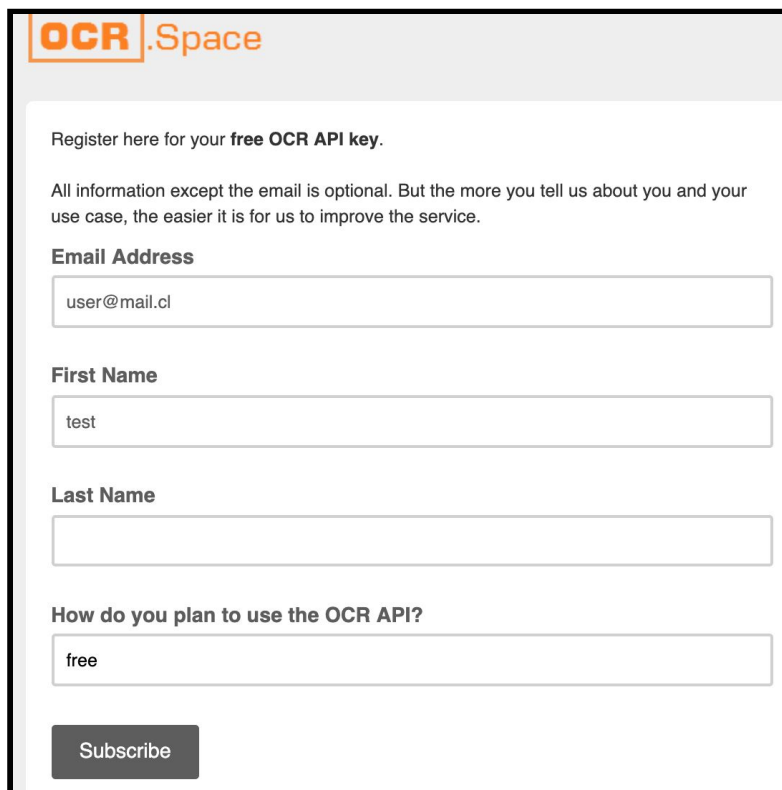
MÓDULO OCR SPACE

Esta API te proporciona OCR para documento de forma gratuita con máximo de tamaño de 1MB por archivo.

Se debe ingresar a la pagina <https://ocr.space> , crear una cuenta y completar los datos del comando.

Esto devolverá la información del documento en una variable lista para usar con Rocketbot Studio.

1. Ir a <https://ocr.space> para registrarse



The image shows a registration form for OCR.Space. At the top, the logo "OCR.Space" is displayed in orange. Below the logo, the text "Register here for your free OCR API key." is shown. A paragraph follows: "All information except the email is optional. But the more you tell us about you and your use case, the easier it is for us to improve the service." The form contains four input fields: "Email Address" with the value "user@mail.cl", "First Name" with the value "test", "Last Name" (empty), and "How do you plan to use the OCR API?" with the value "free". At the bottom, there is a dark grey button labeled "Subscribe".

OCR.Space

Register here for your **free OCR API key**.

All information except the email is optional. But the more you tell us about you and your use case, the easier it is for us to improve the service.

Email Address

user@mail.cl

First Name

test

Last Name

How do you plan to use the OCR API?

free

Subscribe

- Nos enviarán un mail para confirmar el registro

Next step: Please confirm your email address.

Thank you for using the free OCR.space OCR API:

To complete the OCR API sign-up, **please click the link in the email we just sent you**.

Once that is done, you will automatically receive a second email with your personal API key.

- Click en el botón del mail:

Please confirm your email to receive your personal API key.

Yes, subscribe me to this list.

If you received this email by mistake, simply delete it. You won't be subscribed if you don't click the confirmation link above.

- Nos enviarán otro mail con la API Key:

OCR.Space

Hello **Booket**,

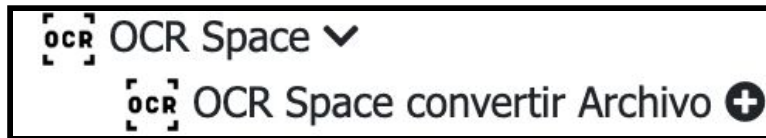
Your free OCR API key is:

69-610-00057

Thank you for using the free OCR API. You can check the API uptime and performance at the status page <https://status.ocr.space>.

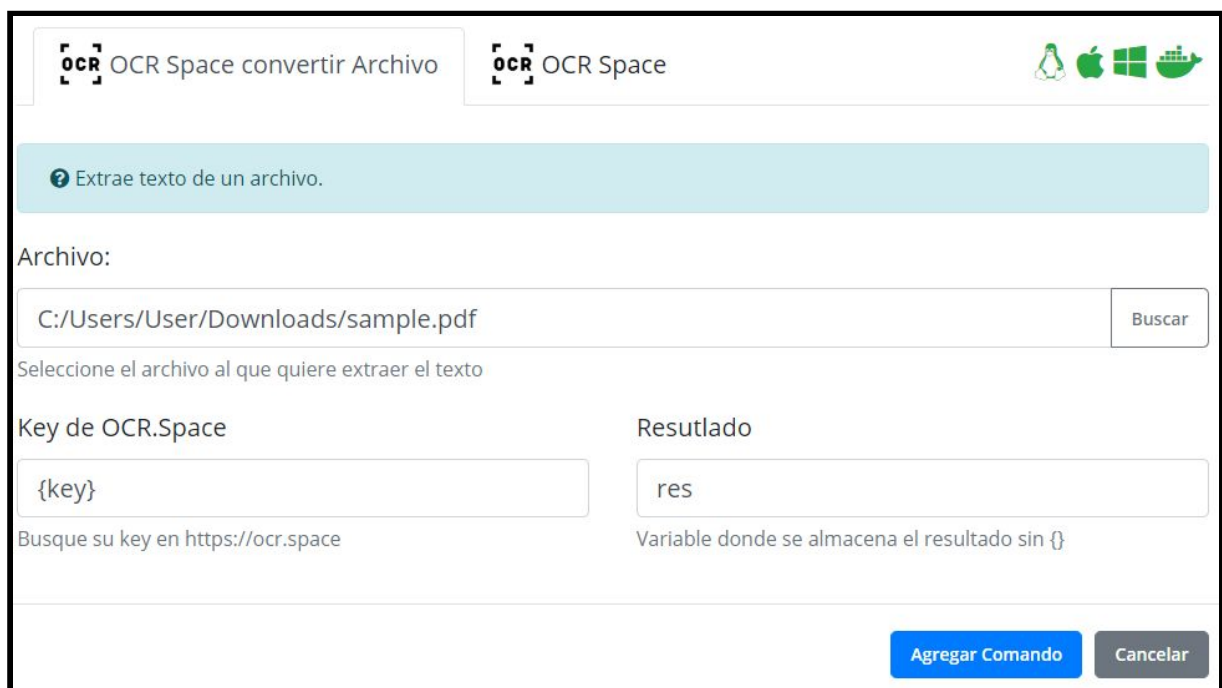
Enjoy - and help justify keeping it free and current by sharing!

2. Iremos al menú “**Mods**” y abriremos el componente de “**OCR Space**”:



Pondremos los siguientes datos:

- Archivo: ruta al archivo del cual queremos extraer el texto.
- Key de OCR Space: Key que nos llegó al correo luego de registrarnos.
- Resultado: Variable donde retornaremos la data, el nombre de la variable va sin llaves “{}”.



OCR Space convertir Archivo OCR Space

Extrae texto de un archivo.

Archivo:

C:/Users/User/Downloads/sample.pdf Buscar

Seleccione el archivo al que quiere extraer el texto

Key de OCR.Space Resutlado


{key} res

Busque su key en <https://ocr.space> Variable donde se almacena el resultado sin {}

Agregar Comando Cancelar


3. Ejecutamos el comando y revisamos la variable, obtendremos algo como esto:


Resultado de variable:

1		res	{'ParsedResults': [{'TextOverlay': {'Lines': [{'LineText': 'A Simple PDF File', 'Words': [{'WordText': 'A', 'Left': 86.0, 'Top': 67.0, 'Height': 26.0, 'Width': 25.0}, {'WordText': 'Simple', 'Left': 122.0, 'Top': 67.0, 'Height': 33.0, 'Width': 107.0}, {'WordText': 'PDF', 'Left': 243.0, 'Top': 67.0, 'Height': 26.0, 'Width': 68.0}, {'WordText': 'File', 'Left': 325.0, 'Top': 67.0, 'Height': 26.0, 'Width': 54.0}], 'MaxHeight': 33.0, 'MinTop': 67.0}, {'LineText': 'This is a small demonstration _pdf file -', 'Words': [{'WordText': 'This', 'Left': 96.0, 'Top': 128.0, 'Height': 10.0, 'Width': 25.0}, {'WordText': 'is', 'Left': 125.0, 'Top':
---	---	-----	--

4. Para limpiar los datos debemos iterar por cada página del PDF:

Debemos iterar con For:

 For

 El cuerpo del bucle se ejecuta tantas veces como elementos tenga el elemento iterable

Variables

Elemento iterable

pdf


{res}["ParsedResults"]


ej: [variable] > 8 | {variable} == 1

Agregar Comando

Cancelar

Luego dentro del **For** debemos utilizar “**Asignar Variable**” para dejar sólo el texto de cada página:

 Asignar Variable

 Puede asignar un dato a una variable o un comando predefinido.

Dato:

{pdf}["ParsedText"]

Asignar resultado a variable

{pdf}


Agregar Comando


Cancelar

La variable {pdf} irá quedando así:

pdf	Simple PDF File 2 .continued from page 1. Yet more text And more text. And more text And m Oh, how boring typing this stuff. But not as boring as watching paint dry. And more text And n text The end: and just as well.
-----	---

Finalmente, a modo de prueba, podemos agregar el comando “**Alerta**” para ir mostrando los datos

 **Alerta**

 Muestra un cartel de mensaje



Mensaje


{pdf}


Agregar Comando



Cancelar




**La secuencia completa se verá así:


  Descripción: 




 OCR Space convertir Archivo
`{"module_name":"OCRSpace","module":"GetOCR","File":"C:/Users/User/Downloads/sample.pdf","result":"res","key":"{key}"}`


  Descripción: Iterar sobre todas las hojas que tiene el PDF 

 For
Estado Elemento iterable pdf `{"iterable":"{res}[\\"ParsedResults\\"]","count":2}`
Res: *false*


  Descripción: Dejar solo el texto 

 Asignar Variable
`pdf = {pdf}["ParsedText"]`

  Descripción: Alert para validar el texto 

 Alerta
Mensaje {pdf}