
	IMAGE_READER<Nombre Proyecto> Manual de Usuario	Agosto, 2023
---	--	---------------------



IMAGE_READER
Manual de Usuario

	<p>IMAGE_READER<Nombre Proyecto></p> <p>Manual de Usuario</p>	<p>Agosto, 2023</p>
---	---	----------------------------

Versión: 0100


Fecha: 08/2023

[Versión del Producto]


Queda prohibido cualquier tipo de explotación y, en particular, la reproducción, distribución, comunicación pública y/o transformación, total o parcial, por cualquier medio, de este documento sin el previo consentimiento expreso y por escrito de TECHANT SA.

HOJA DE CONTROL

Proyecto	<Nombre Proyecto>		
Entregable	Manual de Usuario		
Autor	<Nombre de la Empresa>		
Versión/Edición	0100	Fecha Versión	14/08/2023
Aprobado por		Fecha Aprobación	17/08/2023
		Nº Total de Páginas	13

	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---

1	DESCRIPCIÓN DEL SISTEMA	4
1.1	Objeto.....	4
1.2	Alcance	4
1.3	Funcionalidad.....	4
2	MAPA DEL SISTEMA	5
2.1	Modelo Lógico.....	5
2.2	Navegación	5
2.2.1	Ingresa a input:	6
2.2.2	Ingresa a output:	9
3	DESCRIPCIÓN DEL SISTEMA	12
3.1.1	Mensajes de error.....	12
4	FAQ	13

	<p>IMAGE_READER<Nombre Proyecto></p> <p>Manual de Usuario</p>	<p>Agosto, 2023</p>
---	---	----------------------------

1 DESCRIPCIÓN DEL SISTEMA

1.1 Objeto

El documento se entrega con el fin de instruir la utilización del programa ejecutable “IMAGE_READER”, no pretende servir como guía de distribución ni soporte técnico de la misma.

1.2 Alcance

Se espera que únicamente el usuario final sea quien use este documento como guía.

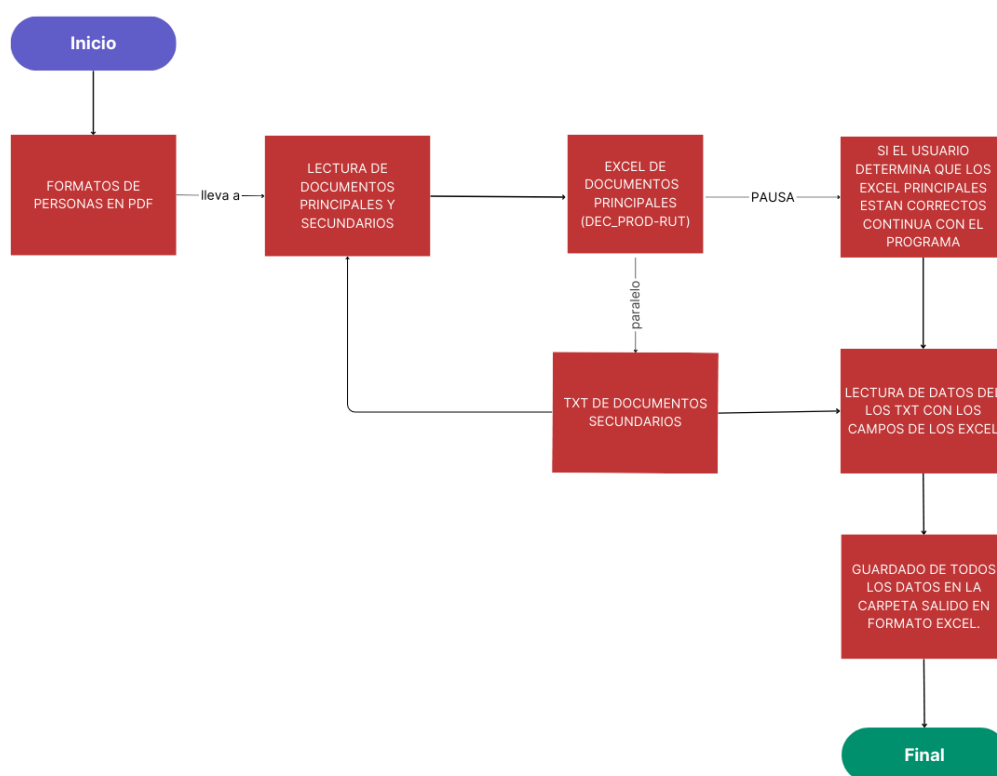
1.3 Funcionalidad

El programa ejecutable “IMAGE_READER” tiene como propósito procesar los documentos correspondientes a diferentes personas con necesidad de identificar si cuentan con todos los formatos y consignando la información de los ya nombrados en un archivo Excel con tablas rotuladas.

2 MAPA DEL SISTEMA

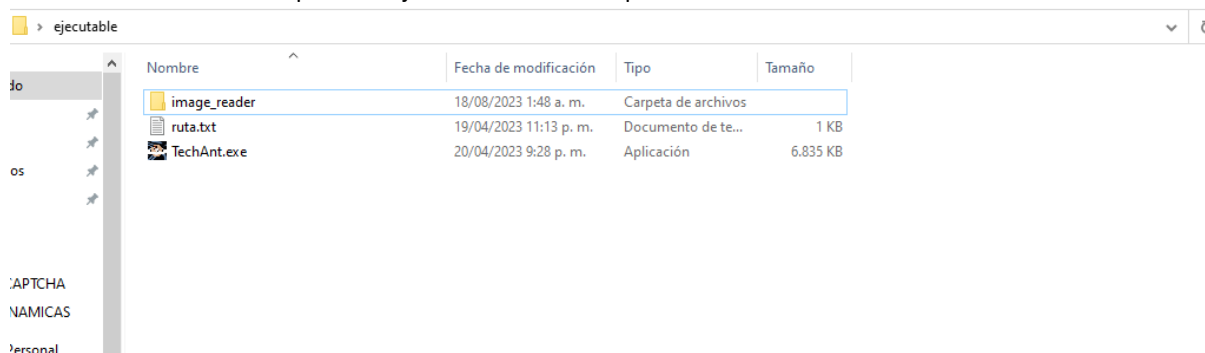
2.1 Modelo Lógico


IMAGE_READER



2.2 Navegación

A la hora de entrar a la carpeta del ejecutable se tiene la primera vista:



	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---


Debe ingresar a la carpeta image_reader:

Nombre	Fecha de modificación	Tipo	Tamaño
__pycache__	15/08/2023 9:10 p. m.	Carpeta de archivos	
build	18/08/2023 4:08 a. m.	Carpeta de archivos	
input	12/08/2023 10:18 p. m.	Carpeta de archivos	
output	15/08/2023 8:56 p. m.	Carpeta de archivos	
Imagen2Letra.py	29/06/2023 9:48 p. m.	Python File	7 KB
j2.py	29/06/2023 1:41 a. m.	Python File	4 KB
lectoraws.py	16/08/2023 6:58 p. m.	Python File	9 KB
main.py	13/08/2023 12:14 a. m.	Python File	9 KB
Pdf2Png.py	12/08/2023 9:16 p. m.	Python File	3 KB
Readme.txt	18/08/2023 4:09 a. m.	Documento de te...	1 KB
requirements.txt	27/06/2023 5:33 p. m.	Documento de te...	4 KB
setup.py	19/04/2023 10:42 p. m.	Python File	1 KB
tesseract-ocr-w64-setup-5.3.0.20221222.e...	17/02/2023 5:29 p. m.	Aplicación	52.163 KB
writeExcel.py	29/06/2023 9:48 p. m.	Python File	24 KB

2.2.1 Ingresa a input:

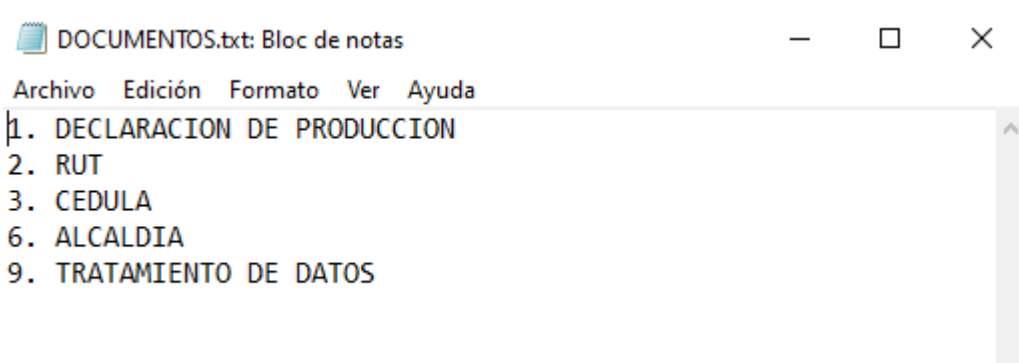
Nombre	Fecha de modificación	Tipo	Tamaño
1. DECLARACION DE PRODUCCION	12/08/2023 10:17 p. m.	Carpeta de archivos	
2. RUT	12/08/2023 10:17 p. m.	Carpeta de archivos	
3. CEDULA	12/08/2023 10:18 p. m.	Carpeta de archivos	
4. CARTA REPRESENTANTE LEGAL	20/04/2023 10:14 p. m.	Carpeta de archivos	
5. DOCUMENTO EQUIVALENTE	24/04/2023 8:52 a. m.	Carpeta de archivos	
6. ALCALDIA	12/08/2023 10:18 p. m.	Carpeta de archivos	
7. FORMATO DE VINCULACION	14/04/2023 6:58 a. m.	Carpeta de archivos	
8. SISBEN	14/04/2023 7:17 a. m.	Carpeta de archivos	
9. TRATAMIENTO DE DATOS	12/08/2023 10:18 p. m.	Carpeta de archivos	
zborrar	15/08/2023 8:53 p. m.	Carpeta de archivos	
DOCUMENTOS.txt	12/08/2023 10:18 p. m.	Documento de te...	1 KB
FORMATO.txt	3/05/2023 11:19 p. m.	Documento de te...	2 KB

En input se encuentra por carpetas, los formatos a revisar se deben colocar correspondientemente en cada una de ellas en formato PDF, es importante respetar la obligatoriedad de los documentos 1. DECLARACION DE PRODUCCION, 2. RUT y 3. CEDULA; y se tiene un archivo 2(DOS) archivos .txt

	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---

1. DOCUMENTOS.txt: aquí se podrá ajustar que documentos se va a analizar, si la carpeta esta creada pero no está asignado en este archivo, el programa no lo leerá.

Ejemplo:



En este caso se están leyendo 5 documentos, debido a que los nombres de la carpeta eran 6. ALCALDIA y 9. TRATAMIENTO DE DATOS se colocan igual, pero se podría editar dichas carpetas para que quedaran con un incremento numérico si así se quiere.

2. FORMATO.txt: Esté archivo contiene indicaciones rápidas para poder ajustar las carpetas obligatorias y opcionales:

```
DOCUMENTOS OBLIGATORIOS:
1. DECLARACION DE PRODUCCION
2. RUT
3. CEDULA

PARA BUSQUEDA AVANZADA:
- Crear un txt en la carpeta que requiera una busqueda avanzada,
  el nombre del txt debe ser "TIPO BUSQUEDA".
- Escribir en el txt el siguiente texto sin comillas: "avanzado".
- Guardar.

COLUMNAS A BUSCAR:


Las columnas disponibles para busqueda son:
1. Nombre      2. Apellido    3. Cc      4. Cantidad    5. Unidad    6. Municipio  7. Fecha    8. Act

Para definir la busqueda:
- Crear un txt en la carpeta que requiera una busqueda avanzada,
  el nombre del txt debe ser "COLUMNAS".
- Escribir en el txt las columnas que desee, por ejemplo:
    1. Nombre
    2. Apellido
    6. Municipio
    7. Fecha
- Guardar.

TENER EN CUENTA:
si al buscar columna se añade '7. Fecha' buscará la primera línea de texto donde haya un año y la reescribira en el excel de salida.
si no se añade '7. Fecha' se debe generar otro txt en la ruta de la carpeta en input que recibira el nombre de 'FECHA'.
```

Se extiende la invitación a leerlo la primera vez que corre este programa, allí podrá encontrar los ajustes necesarios para cada carpeta.

En este punto se debe preparar la carpeta 1. DECLARACION DE PRODUCCION, ya que está funciona por medio de los servicios AWS. A continuación, encontrara los ajustes necesarios:

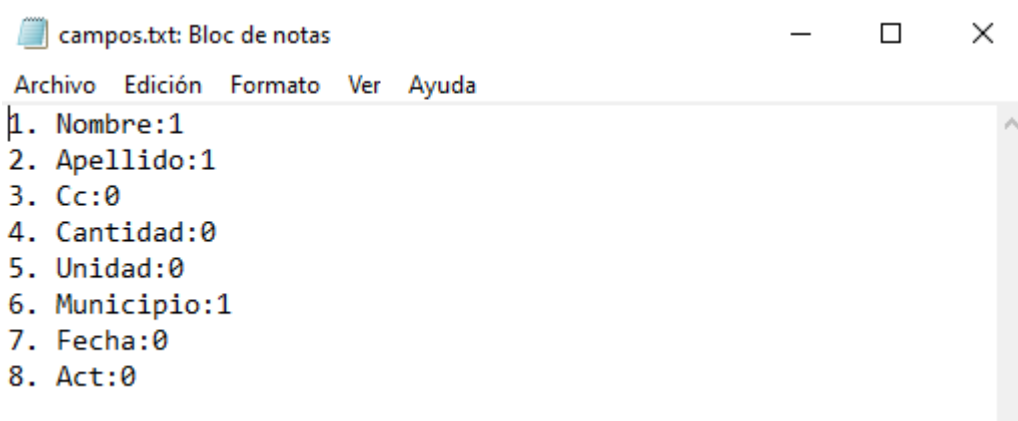
	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---

> Este equipo > Documentos > TECHANT > image_reader > input > 1. DECLARACION DE PRODUCCION				
Nombre	Fecha de modificación	Tipo	Tamaño	
campos.txt	15/08/2023 11:49 p. m.	Documento de te...	1 KB	
ES IM 004 DECLARACION DE PRODUCCION-1-10.pdf	31/03/2023 9:47 p. m.	Microsoft Edge P...	1.893 KB	
GIRO.txt	27/06/2023 1:58 a. m.	Documento de te...	1 KB	

Se puede observar 2(DOS) .txt.

1. campos.txt: aquí define los campos mínimos para que el programa funcione, si el programa no encuentra alguno de estos campos cerrara el programa

Ejemplo:




```

1. Nombre:1
2. Apellido:1
3. Cc:0
4. Cantidad:0
5. Unidad:0
6. Municipio:1
7. Fecha:0
8. Act:0
  
```

En este caso, el programa buscara los campos que se igualaron a 1 como campos mínimos, y los que está en 0 no restringirán el programa.

2. GIRO.txt: Si desea que el pdf quede inclinado 90° a la derecha puede escribir “si” en este txt (Esto puede hacer tardar más el programa).

Una vez realizado estos pasos se debe verificar la carpeta “zborrar” en la ruta “input”:

	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---

Este equipo > Documentos > TECHANT > image_reader > input >					
	Nombre	Fecha de modificación	Tipo	Tamaño	
	1. DECLARACION DE PRODUCCION	18/08/2023 4:21 a. m.	Carpeta de archivos		
	2. RUT	12/08/2023 10:17 p. m.	Carpeta de archivos		
	3. CEDULA	12/08/2023 10:18 p. m.	Carpeta de archivos		
	4. CARTA REPRESENTANTE LEGAL	20/04/2023 10:14 p. m.	Carpeta de archivos		
	5. DOCUMENTO EQUIVALENTE	24/04/2023 8:52 a. m.	Carpeta de archivos		
	6. ALCALDIA	12/08/2023 10:18 p. m.	Carpeta de archivos		
	7. FORMATO DE VINCULACION	14/04/2023 6:58 a. m.	Carpeta de archivos		
	8. SISBEN	14/04/2023 7:17 a. m.	Carpeta de archivos		
	9. TRATAMIENTO DE DATOS	12/08/2023 10:18 p. m.	Carpeta de archivos		
	zborrar	15/08/2023 8:53 p. m.	Carpeta de archivos		
	DOCUMENTOS.txt	12/08/2023 10:18 p. m.	Documento de te...	1 KB	
	FORMATO.txt	3/05/2023 11:19 p. m.	Documento de te...	2 KB	

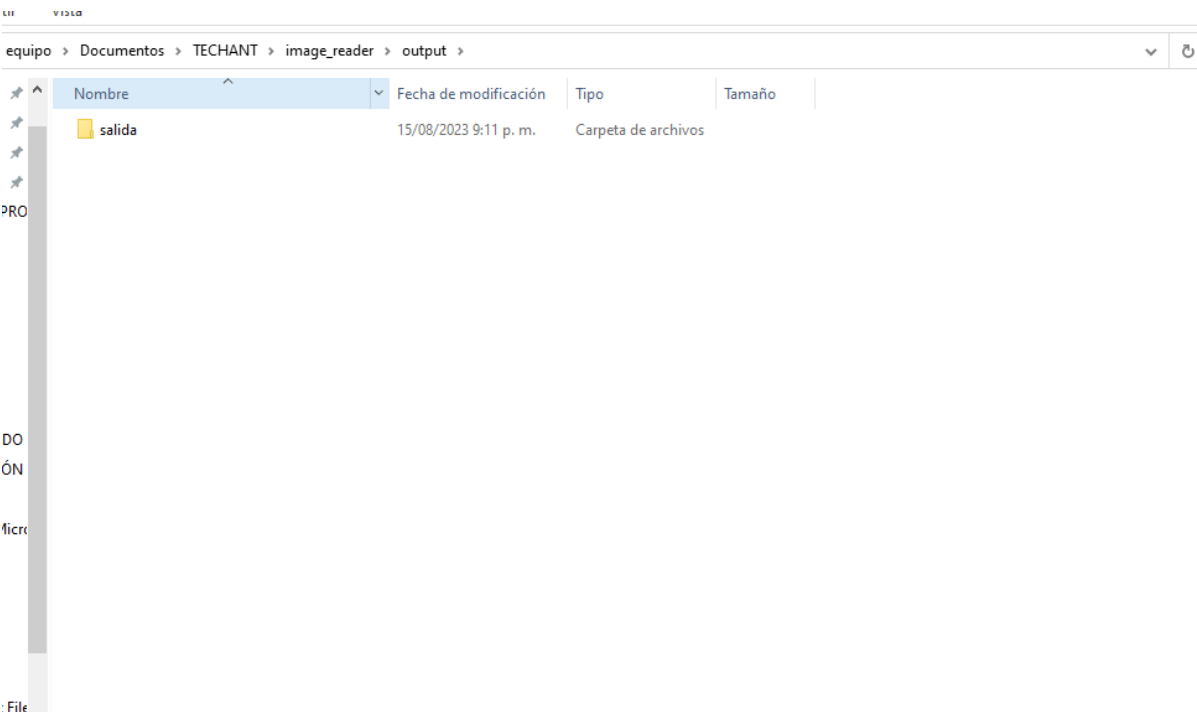
En dicha carpeta se debe asegurar de que solo haya 2(DOS) subcarpetas llamadas “borrar1” y “borrar2” cada una de ellas con una subcarpeta adentro llamada “Procesado” y todo debe estar vacío.

2.2.2 Ingresa a output:

Una vez definidos los parámetros de entrada (input) se debe definir las condiciones de salida (output)

Este equipo > Documentos > TECHANT > image_reader >					
	Nombre	Fecha de modificación	Tipo	Tamaño	
	pycache	15/08/2023 9:10 p. m.	Carpeta de archivos		
	build	18/08/2023 4:10 a. m.	Carpeta de archivos		
	input	18/08/2023 4:21 a. m.	Carpeta de archivos		
	output	15/08/2023 8:56 p. m.	Carpeta de archivos		
	Imagen2Letra.py	29/06/2023 9:48 p. m.	Python File	7 KB	
	j2.py	29/06/2023 1:41 a. m.	Python File	4 KB	
	lectoraws.py	16/08/2023 6:58 p. m.	Python File	9 KB	
	main.py	13/08/2023 12:14 a. m.	Python File	9 KB	
	Pdf2Png.py	12/08/2023 9:16 p. m.	Python File	3 KB	
	Readme.txt	18/08/2023 4:09 a. m.	Documento de te...	1 KB	
	requirements.txt	27/06/2023 5:33 p. m.	Documento de te...	4 KB	
	setup.py	19/04/2023 10:42 p. m.	Python File	1 KB	
	tesseract-ocr-w64-setup-5.3.0.20221222.e...	17/02/2023 5:29 p. m.	Aplicación	52.163 KB	
	writeExcel.py	29/06/2023 9:48 p. m.	Python File	24 KB	


Al revisarla solo debe estar una carpeta llamada “salida”:




Al ejecutar el programa y si todo está funcionando adecuadamente, se debería así:

Nombre	Fecha de modificación	Tipo	Tamaño
2. RUT	15/08/2023 9:11 p. m.	Carpeta de archivos	
3. CEDULA	15/08/2023 9:13 p. m.	Carpeta de archivos	
6. ALCALDIA	15/08/2023 9:13 p. m.	Carpeta de archivos	
9. TRATAMIENTO DE DATOS	15/08/2023 9:13 p. m.	Carpeta de archivos	
salida	15/08/2023 9:11 p. m.	Carpeta de archivos	

io > Documentos > TECHANT > image_reader > output > salida

Nombre	Fecha de modificación	Tipo	Tamaño
 1. DECLARACION DE PRODUCCION_15_0...	15/08/2023 9:12 p. m.	Hoja de cálculo d...	6 KB

Página 10 de 13

	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---


Una vez revisado esto se podrá continuar el programa.

```

PROCESADO: borrar1 , 4.jpg
PROCESADO: borrar1 , 5.jpg
PROCESADO: borrar1 , 6.jpg
PROCESADO: borrar1 , 7.jpg
PROCESADO: borrar1 , 8.jpg
PROCESADO: borrar1 , 9.jpg
PROCESADO: borrar2 , 1.jpg
PROCESADO: borrar2 , 10.jpg
PROCESADO: borrar2 , 2.jpg
PROCESADO: borrar2 , 3.jpg
PROCESADO: borrar2 , 4.jpg
PROCESADO: borrar2 , 5.jpg
PROCESADO: borrar2 , 6.jpg
PROCESADO: borrar2 , 7.jpg
PROCESADO: borrar2 , 8.jpg
PROCESADO: borrar2 , 9.jpg
Presiona una tecla para continuar...

```

Al continuar se leerá todos los documentos y se obtiene un Excel de cada formato que se guardara en la carpeta salida, y un Excel llamado salidaFECHADEEJECUCION.xlsx en la carpeta “output”.

	<p align="center">IMAGE_READER<Nombre Proyecto></p> <p align="center">Manual de Usuario</p>	<p align="center">Agosto, 2023</p>
---	---	---

3 DESCRIPCIÓN DEL SISTEMA

El programa funciona por consola, por lo que no se tiene interfaz para dar detalles

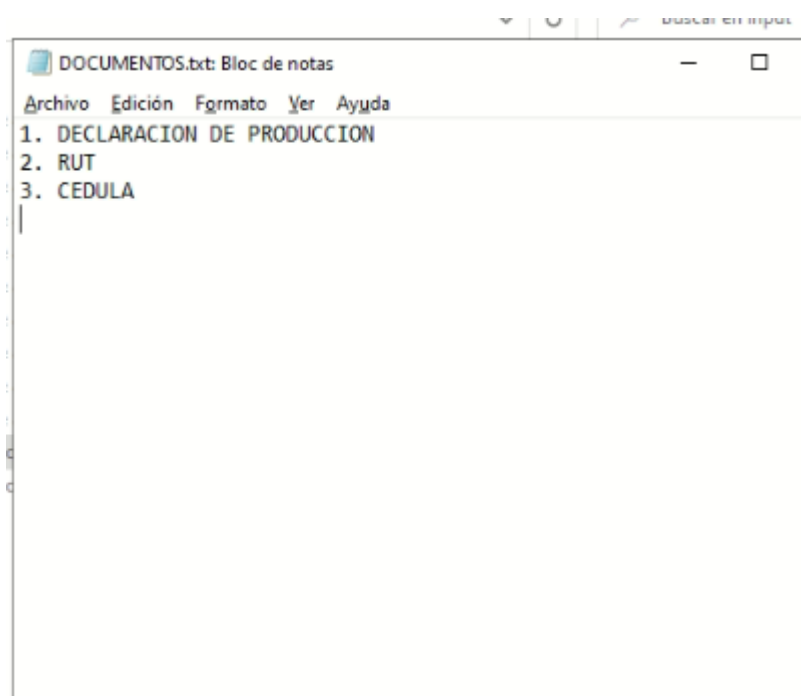
3.1.1 Mensajes de error

Un error típico sucede por no revisar que las carpetas en el output estén vacías, esto puede dañar los Excel de salida.


Si AWS no puede encontrar los campos mínimos en la primera página del formato “1. DECLARACION DE PRODUCCION” el programa se cerrará.

Otro error típico se puede encontrar al dejar espacios en blanco en el bloc de notas.

Ejemplo:



En casos como esté el programa buscara una carpeta llamada “ ” e intentar hacer el proceso de extracción de imágenes

	<p>IMAGE_READER<Nombre Proyecto></p> <p>Manual de Usuario</p>	<p>Agosto, 2023</p>
---	---	----------------------------

4 FAQ

¿Cuántas paginas puede tener los formatos?

La prueba mas extensa hasta el momento fue de 300 páginas por formato.

¿Qué sucede si los formatos tienen mas o menos paginas entre ellos?

Deben coincidir las paginas ya que el rango de lectura hará que falle si son menores.

¿Qué consumo tiene?

Actualmente está utilizando 12 procesadores en paralelo para lograr más eficiencia.