



Ptéryx

Índice

1	– Introducción	1
2	– Instalación y Requisitos del Sistema	2
2.1	– Requisitos Mínimos	2
2.2	– Instalación	2
2.3	– Idiomas	2
3	– Interfaz de Usuario	3
3.1	– Pre-registro	3
3.2	– Adquisición de la Licencia	4
3.3	– Post-registro – Pantalla de Inicio	4
4	– Módulo de Preprocesamiento de PDF	4
4.1	– Organización de Archivos	4
4.2	– Procesamiento	6
4.3	– Ejecución de OCR	7
4.4	– Finalización	8
4.5	– Proceder	8
5	– Módulo de Búsqueda	9
5.1	– Búsqueda por Especie	9
5.2	– Búsqueda por Lista de Especies	10
5.3	– Búsqueda por Términos o Frases	12
6	– Exportación y generación de informes.....	13
7	– Solución de Problemas	14
8	– Contacto y Soporte	14
9	– Glosario	14
10	– Guía Rápida — Pteryx	15



1. Introducción

La Evolución de la Investigación Científica y la Necesidad de Automatización

Hasta hace poco, la investigación científica dependía de documentos físicos: los investigadores debían hojear páginas, anotar manualmente información relevante y manejar grandes volúmenes de papel. Este proceso consumía mucho tiempo, era laborioso y propenso a errores humanos.

La introducción del formato PDF en 1993 revolucionó la manera en que almacenamos y accedemos a la información. Los documentos comenzaron a archivar digitalmente, simplificando la organización y eliminando pilas de papel. Sin embargo, incluso con este cambio digital, buscar manualmente entre cientos o miles de PDFs sigue siendo una tarea extremadamente tediosa.

Hoy en día, la mayoría de los artículos y libros académicos están disponibles en línea, reduciendo drásticamente la circulación de materiales físicos. En Brasil, el impacto es evidente: en la última década, el 17% de los quioscos y el 29% de las librerías han cerrado, y entre 2015 y 2020, se perdieron al menos 764 bibliotecas públicas (Sistema Nacional de Bibliotecas Públicas). Esto obliga a los investigadores a depender de dispositivos electrónicos para la lectura, lo que, con el tiempo, puede causar fatiga visual y otros problemas oculares.

Los Desafíos de la Búsqueda Manual en PDFs

Buscar manualmente en documentos digitales implica abrir archivos, leer cada página, tomar notas y repetir el proceso continuamente. Este método, además de consumir mucho tiempo, aumenta la probabilidad de errores, ya que la fatiga de la lectura prolongada puede provocar la omisión de datos relevantes. Los estudios indican que leer en pantallas retroiluminadas requiere un mayor esfuerzo visual, impactando en la precisión de los análisis y haciendo el proceso menos eficiente.

Más allá del agotamiento cognitivo, la repetición continua de estas tareas agota física y mentalmente al investigador. Considerando que leer un PDF de 5 páginas toma un promedio de 5 minutos, un profesional necesitaría 8 horas para revisar aproximadamente 100 documentos. Para examinar 3,000 archivos, se requerirían aproximadamente 30 días de trabajo ininterrumpido. ¿Por qué realizar esta tarea manualmente cuando el software puede ejecutarla en pocos minutos, con mayor precisión y sin interrupciones?

Limitaciones de la Búsqueda de Windows y la Solución de Pteryx

Microsoft Windows permite buscar información dentro de múltiples archivos, pero esta funcionalidad tiene limitaciones: solo permite búsquedas de un término a la vez, no genera informes detallados ni organiza los archivos automáticamente. Esto hace que el cribado manual sea lento e ineficaz para la investigación científica.

Pteryx fue desarrollado para resolver este problema. Creado por un investigador en zoología, ofrece una solución robusta y automatizada para analizar documentos científicos en formato PDF. El software preprocesa los archivos, realiza búsquedas automatizadas y genera informes detallados, indicando dónde se encontró cada información y en qué contexto. Esto hace que la revisión de documentos sea más rápida, precisa y eficiente, permitiendo a los investigadores centrarse en el análisis de los datos en lugar de perder tiempo en tareas repetitivas.

Automatizar esta etapa es esencial para mantener el ritmo de la creciente digitalización de la ciencia y garantizar que la investigación continúe avanzando sin los obstáculos del trabajo manual.



2. Instalación y Requisitos del Sistema

2.1 Requisitos Mínimos

- **Sistema Operativo:** Windows 10 o superior
- **Procesador:** Intel Core i5 o equivalente
- **Memoria RAM:** 8GB o más
- **Espacio en Disco:** 2GB libres para la instalación
- **Resolución mínima:** 1280 x 720
- **Software requerido:** Microsoft Excel o equivalente instalado (para la manipulación de informes)
- **Bibliotecas requeridas:** el software incluye todas las dependencias en el ejecutable, sin necesidad de instalación manual.

2.2 Instalación

1. Descargue el instalador de Pteryx desde el sitio web de Pteryx.
2. Ejecute Pteryx_installer.exe, elija la unidad de instalación e introduzca su correo electrónico. El instalador descargará la aplicación y creará la estructura de carpetas necesaria.

2.3 Idiomas

Pteryx detecta automáticamente el idioma de su sistema y se ejecutará en uno de los siguientes: inglés, portugués, español, gallego, búlgaro, francés, ruso, alemán, japonés o mandarín. Para cualquier otro idioma, la interfaz se configurará en inglés por defecto.

3. Interfaz de Usuario

3.1 Pre-registro

Antes de comprar y registrar el programa, funcionará en modo demo, mostrando un botón "registrarse ahora" en la pantalla principal (Fig. 1). En modo demo, el preprocesamiento no funciona y las búsquedas están limitadas a tres en los modos "Búsqueda por especie" y "palabra o frase".

Después de completar las tres búsquedas, el programa lo llevará directamente a la pantalla de registro (Fig. 2).

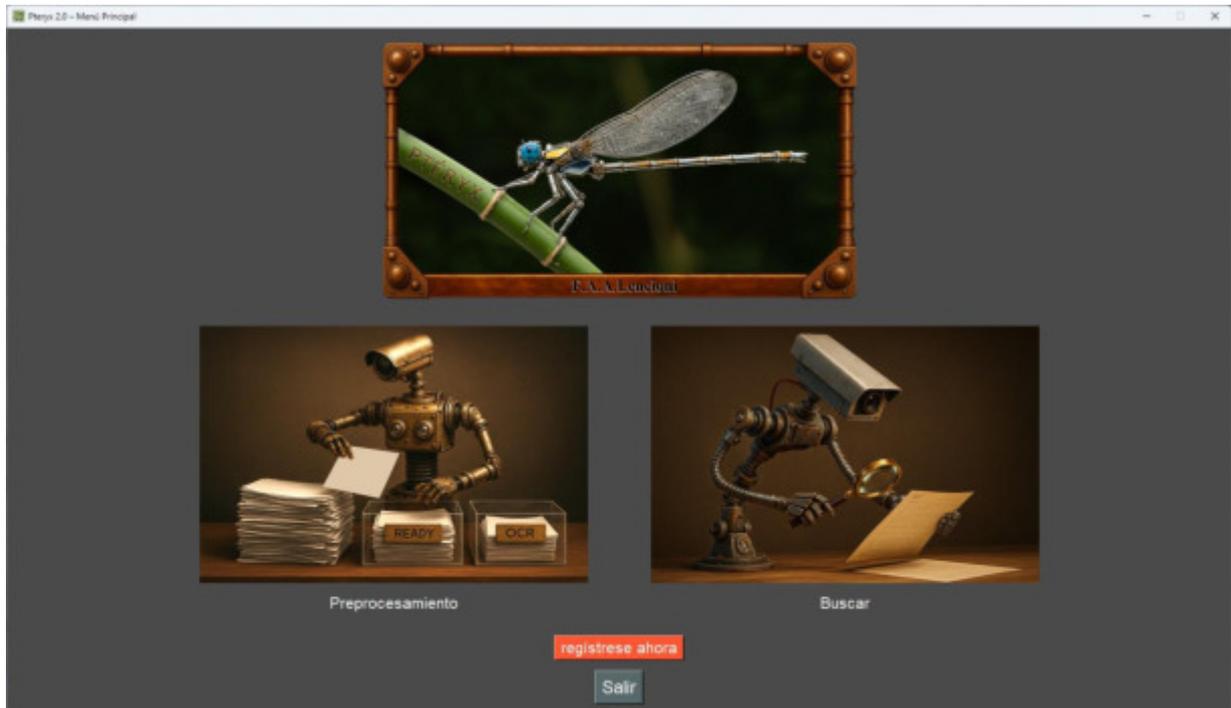


Fig. 1 – Menú principal (pantalla de pre-registro).

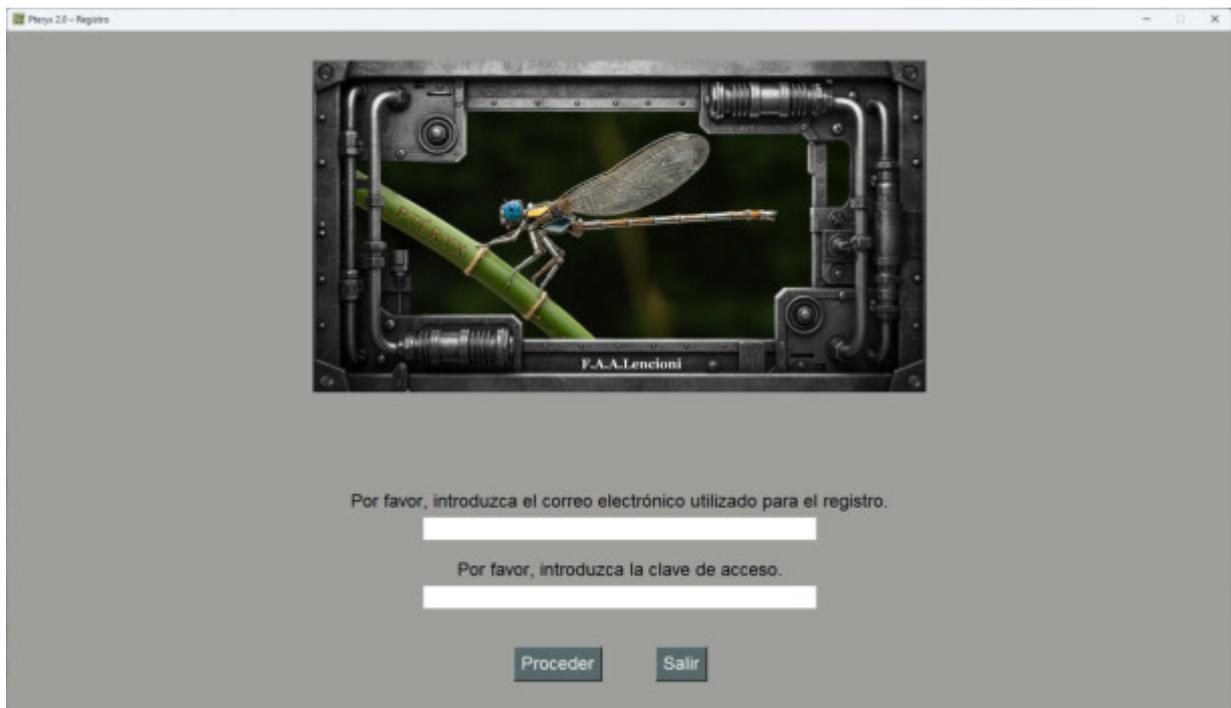


Fig. 2 – Pantalla de registro.

3.2 Aquisição da Licença

Para adquirir a licença, acesse o site www.pteryx.com.br e, assim que o pagamento for efetuado, você receberá no e-mail cadastrado a chave de acesso para desbloquear o software.

3.3 Post-registro – Pantalla de Inicio

Después del registro, accederá a la pantalla principal (Fig. 3). Desde aquí puede seleccionar uno de los dos módulos:

- **Preprocesamiento** (Preprocesamiento de PDFs)
- **Búsqueda** (Módulo de búsqueda)

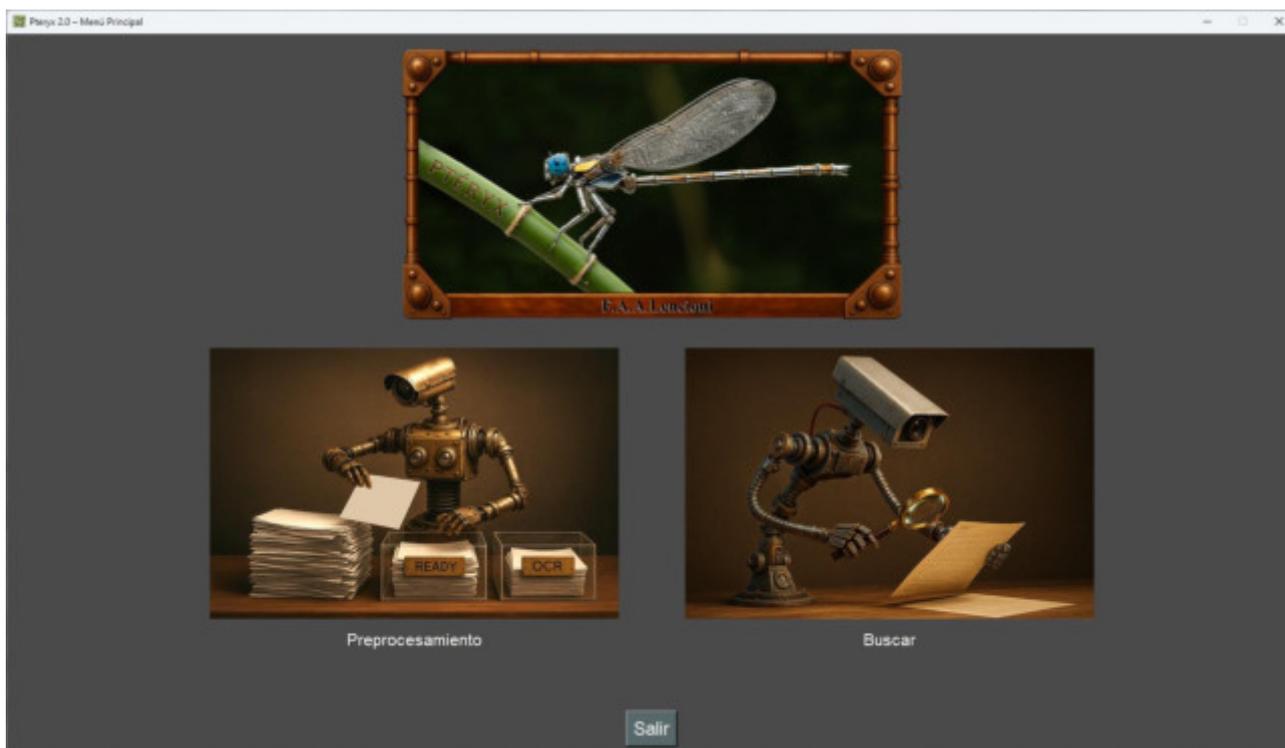


Fig. 3 – Menú principal.

4. Módulo de Preprocesamiento de PDF

El software está diseñado para buscar en PDFs y generar informes detallados de estas búsquedas. Para ello, los PDFs deben ser legibles, buscables, sin errores y organizados en una ubicación predeterminada.

El sistema de preprocesamiento hace esto en dos pasos:

Paso 1: busca y copia todos los PDFs de una carpeta o disco al lugar donde serán procesados (el usuario tiene la opción de colocar manualmente los PDFs en la carpeta **1_PDFs_to_Process**).

Paso 2: examina todos los PDFs, buscando errores, duplicados y PDFs basados en imágenes (que se mueven a la carpeta **2_PDFs_to_OCR**).

4.1 Organización de Archivos

- Haga clic en el botón "Preprocesamiento" en la pantalla principal.
- Puede copiar y pegar manualmente los PDFs en la carpeta **1_PDFs_to_Process** o hacer clic en **"Recolectar PDFs"** y especificar qué carpetas o discos el programa debe recopilar los PDFs.
- El software solo copiará sus archivos; su estructura original se mantendrá intacta.
- Si elige varias carpetas que contienen copias del mismo PDF, se guardarán con el sufijo (1). Si existen más copias, se guardarán con números secuenciales (2, 3, 4, etc.).

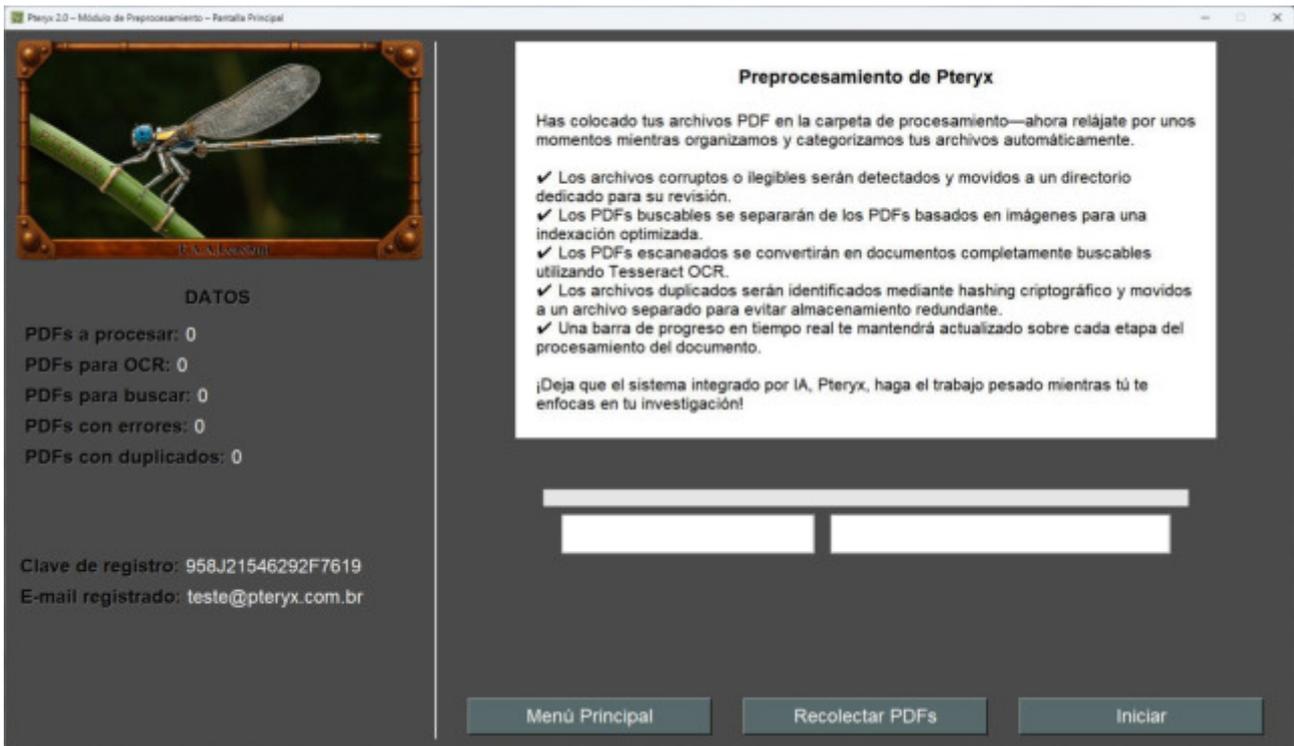


Fig. 4 – Preprocesamiento de PDF.

- Haga clic en el botón "Recolectar PDFs".

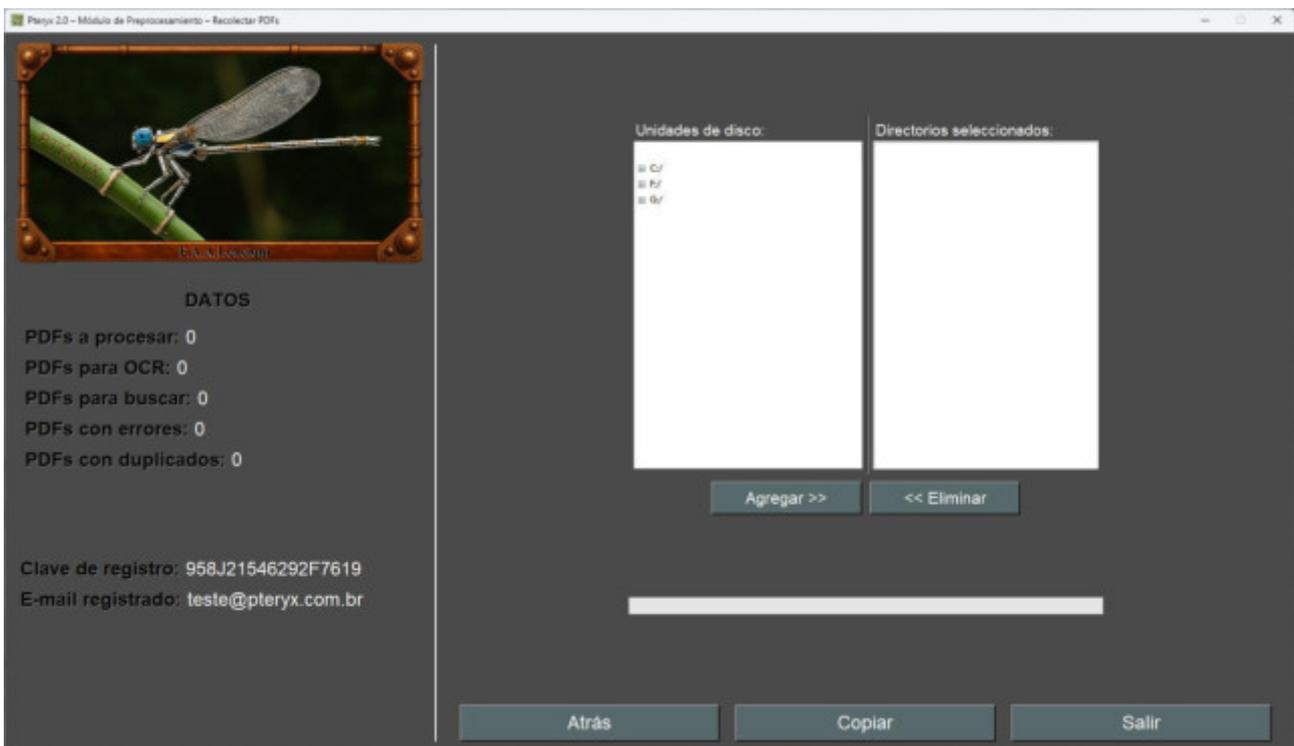


Fig. 5 – Recolectar PDFs, eligiendo carpetas.

- Elija las carpetas o discos de donde se recopilarán los PDFs y añádalos a los “**Directorios seleccionados**”.
- Presione Copiar y todos los archivos PDF de las carpetas o discos seleccionados se copiarán en **1_PDFs_to_Process**.

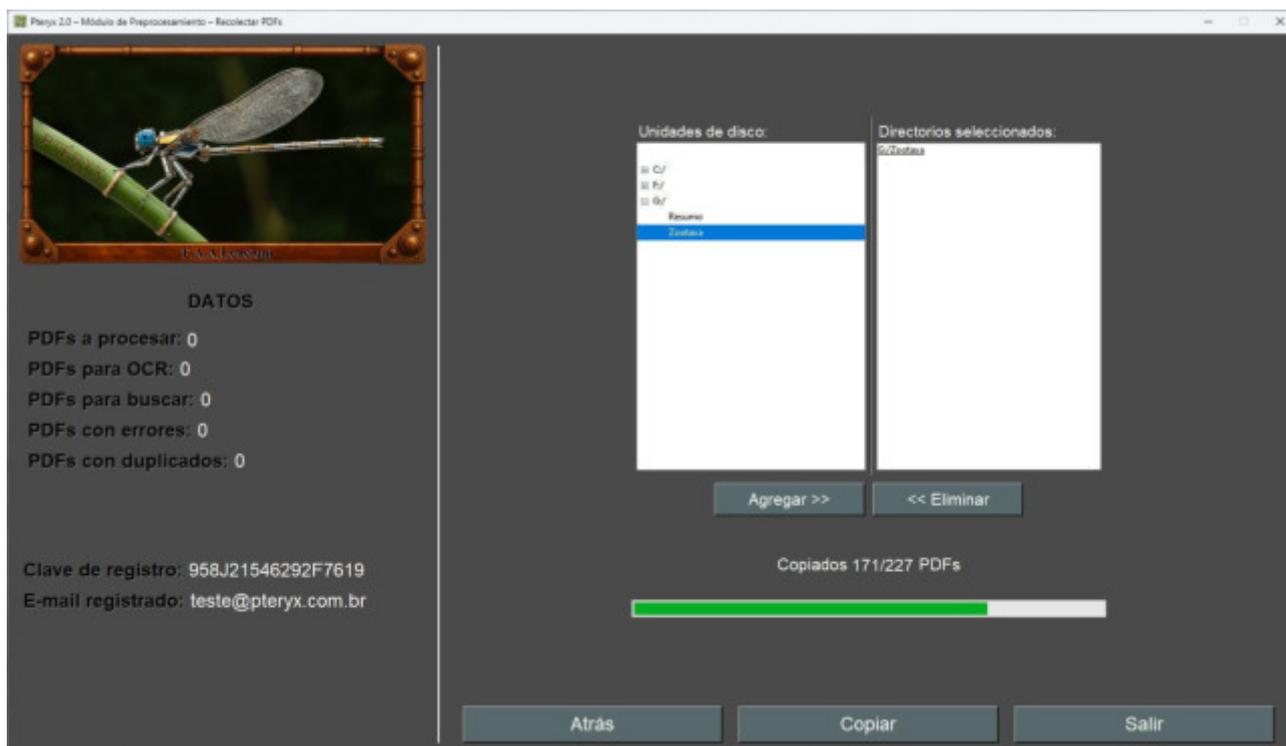


Fig. 6 – Recolectar PDFs, copiando archivos.

4.2 Procesamiento

Una vez que todos los PDFs estén en la carpeta **1_PDFs_to_Process**, regrese al menú principal de preprocesamiento y haga clic en **"Iniciar"**. El sistema procesará todos los archivos, verificando:

- Archivos corruptos o con errores → se moverán a **4_PDFs_with_Errors** (Fig. 7)
- Archivos duplicados → se moverán a **5_PDFs_with_duplicates** (Fig. 8)

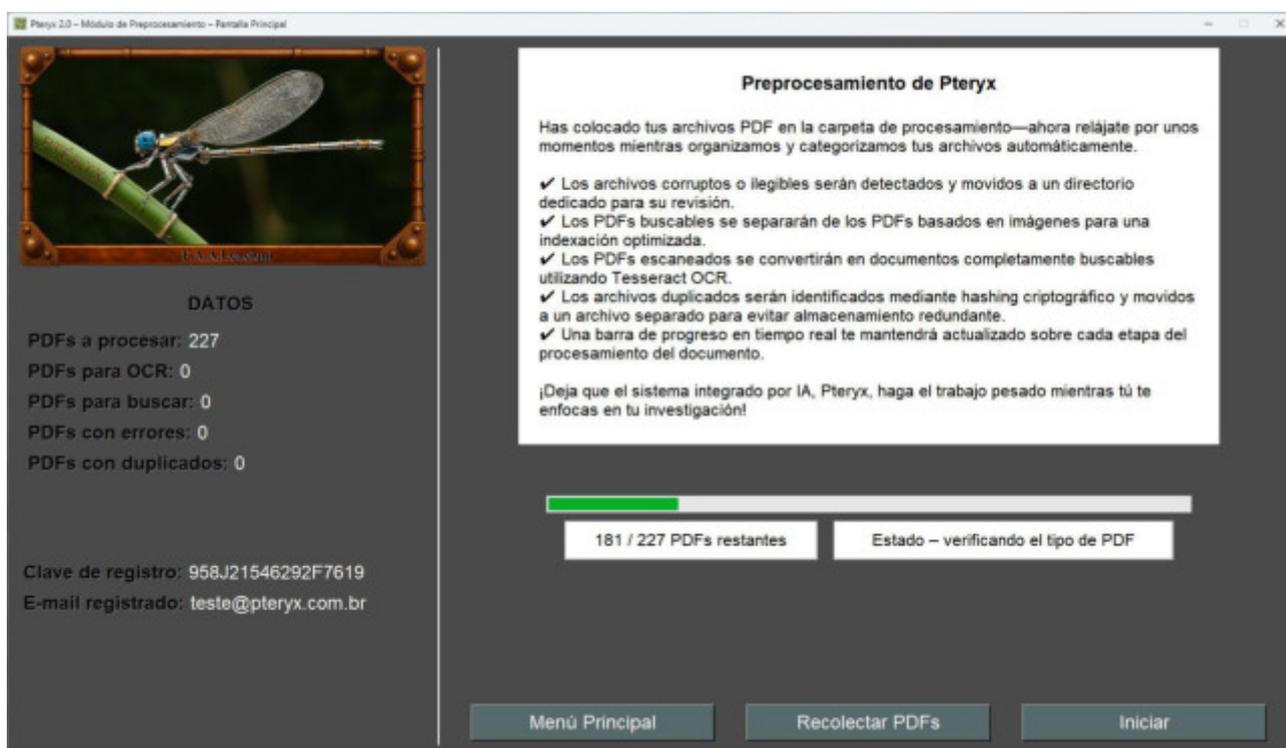


Fig. 7 – Verificación del tipo de archivo.

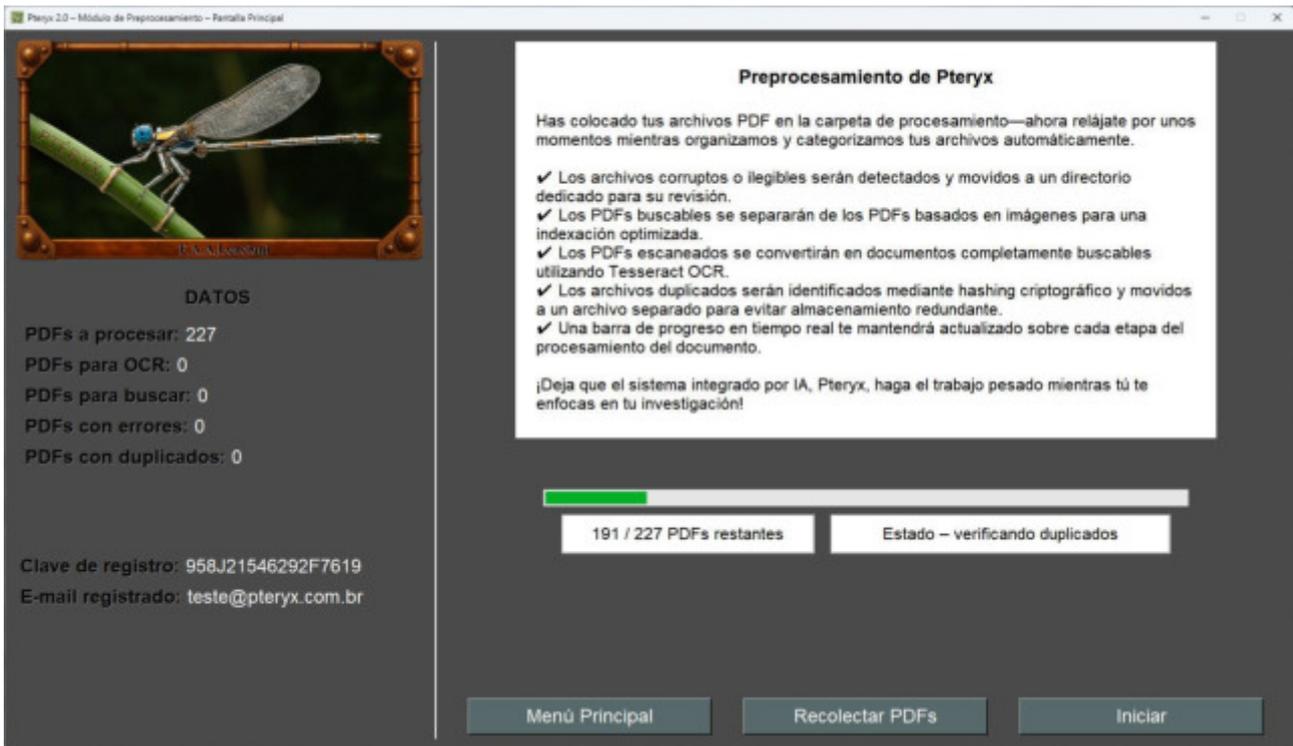


Fig. 8 – Verificación de archivos duplicados.

- PDFs basados en imagen que requieren OCR → se moverán a **2_PDFs_to_OCR**.
- PDFs listos para búsqueda → se moverán a **3_PDFs_to_search**.

4.3 Ejecución de OCR

Una vez finalizado el preprocesamiento, si hay archivos que requieren OCR (PDFs basados en imagen), el sistema mostrará el número de archivos y preguntará si desea ejecutar el OCR (Fig. 9). Si elige “Sí”, simplemente espere a que todos los archivos sean procesados (Fig. 10).

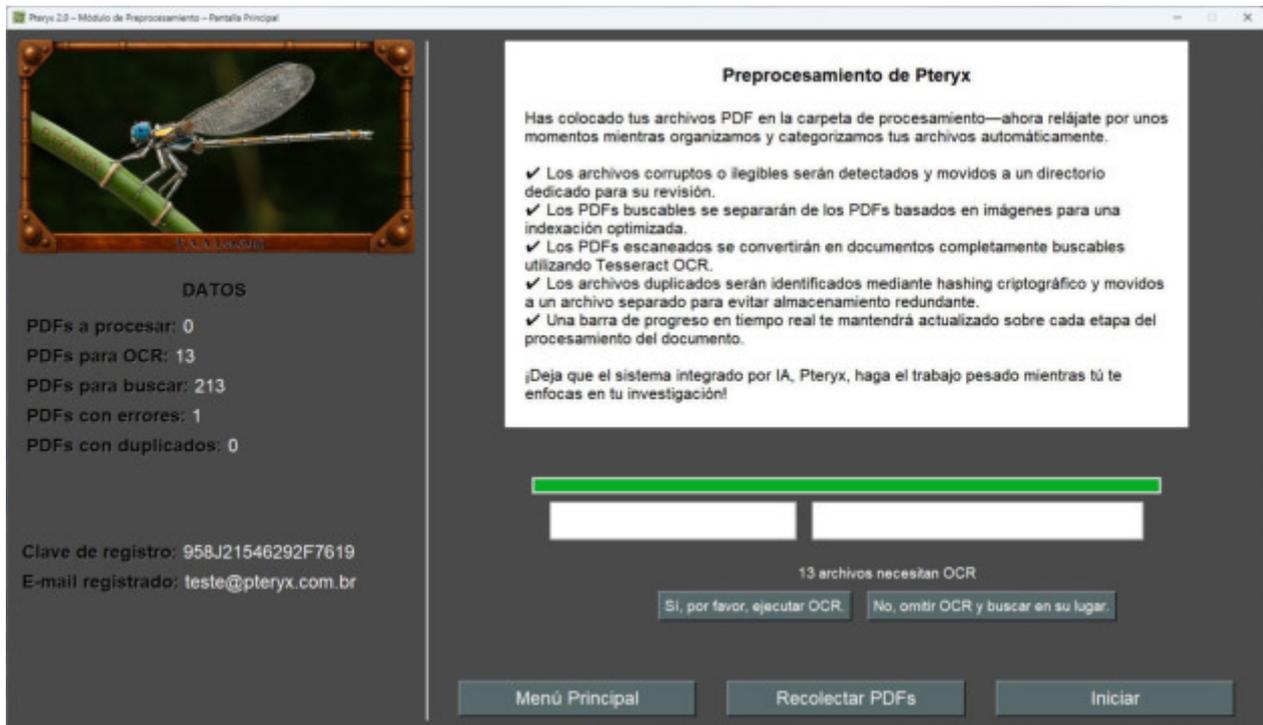


Fig. 9 – Solicitud de ejecución de OCR.

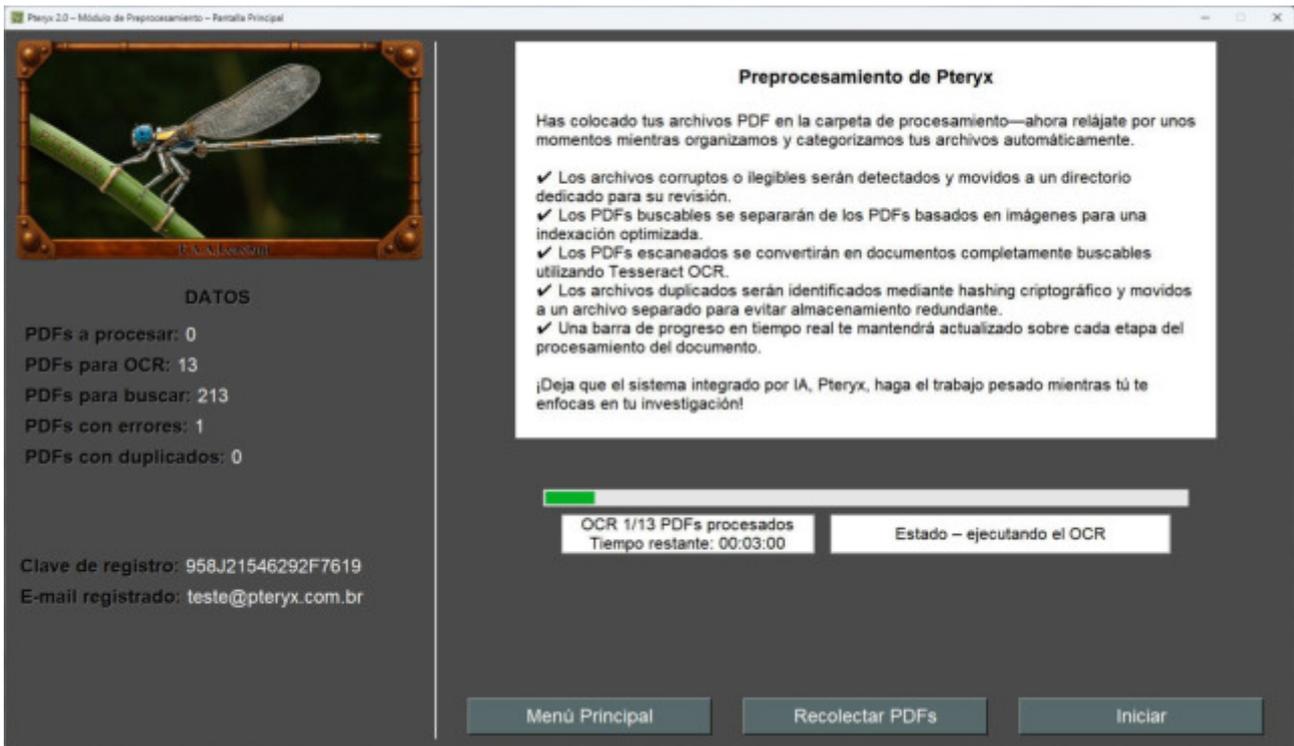


Fig. 10 – Ejecutando OCR.

4.4 Finalización

Después de completar el proceso de OCR, los archivos se moverán a la carpeta **3_PDFs_to_search** y se mostrará un informe de todos los PDFs procesados (Fig. 11).

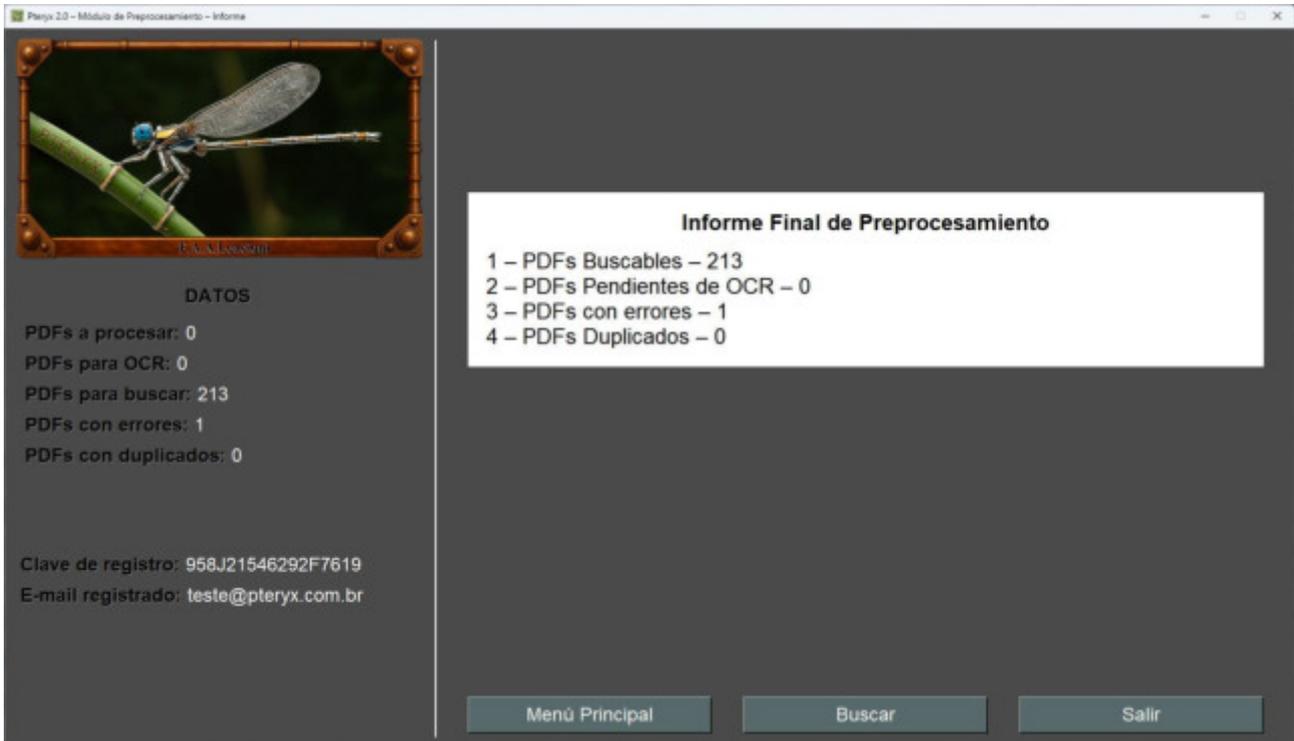


Fig. 11 – Informe final de preprocesamiento.

4.5 Proceder

Ahora, regrese al Menú Principal para realizar sus búsquedas.

5. Módulo de Búsqueda

Este es el módulo principal, que ofrece tres métodos de búsqueda. Cada uno genera informes detallados y fáciles de revisar (Fig. 12).

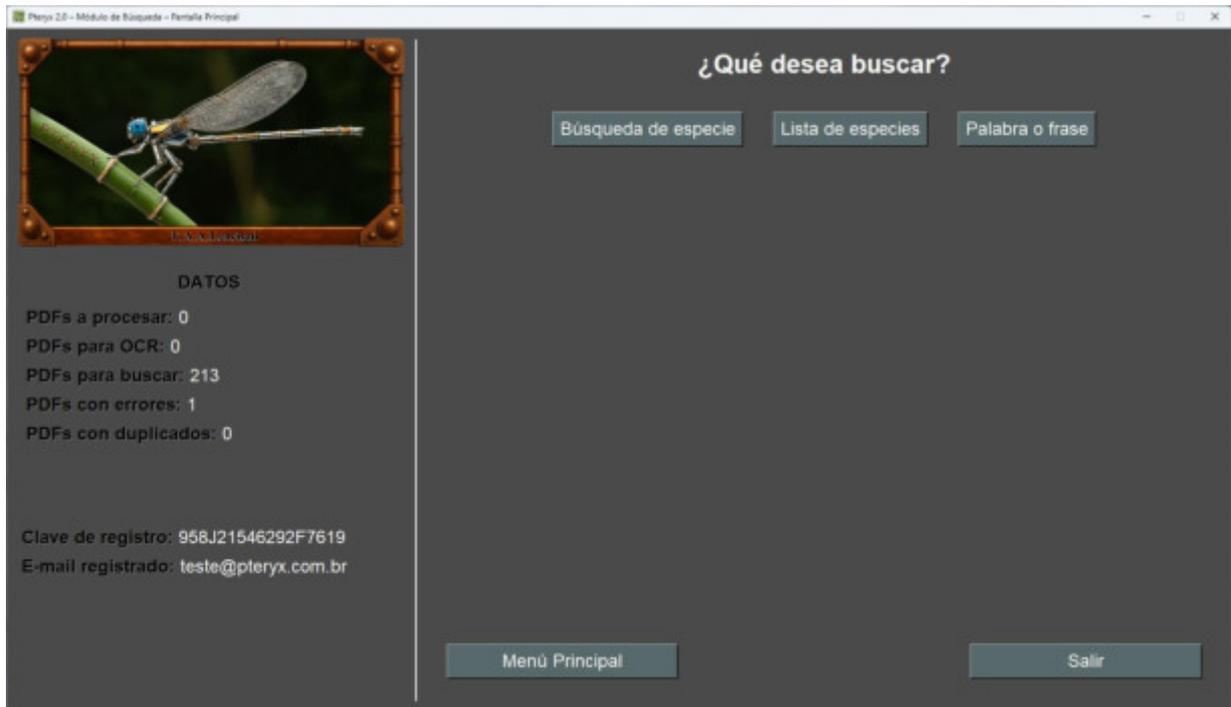


Fig. 12 – Módulo de búsqueda.

5.1 Búsqueda por Especie

Esta función busca todas las ocurrencias de una especie específica en los PDFs almacenados en la carpeta 3_PDFs_to_search. La búsqueda puede realizarse con el ingreso del género (Fig. 13) o sin el ingreso del género (Fig. 14).

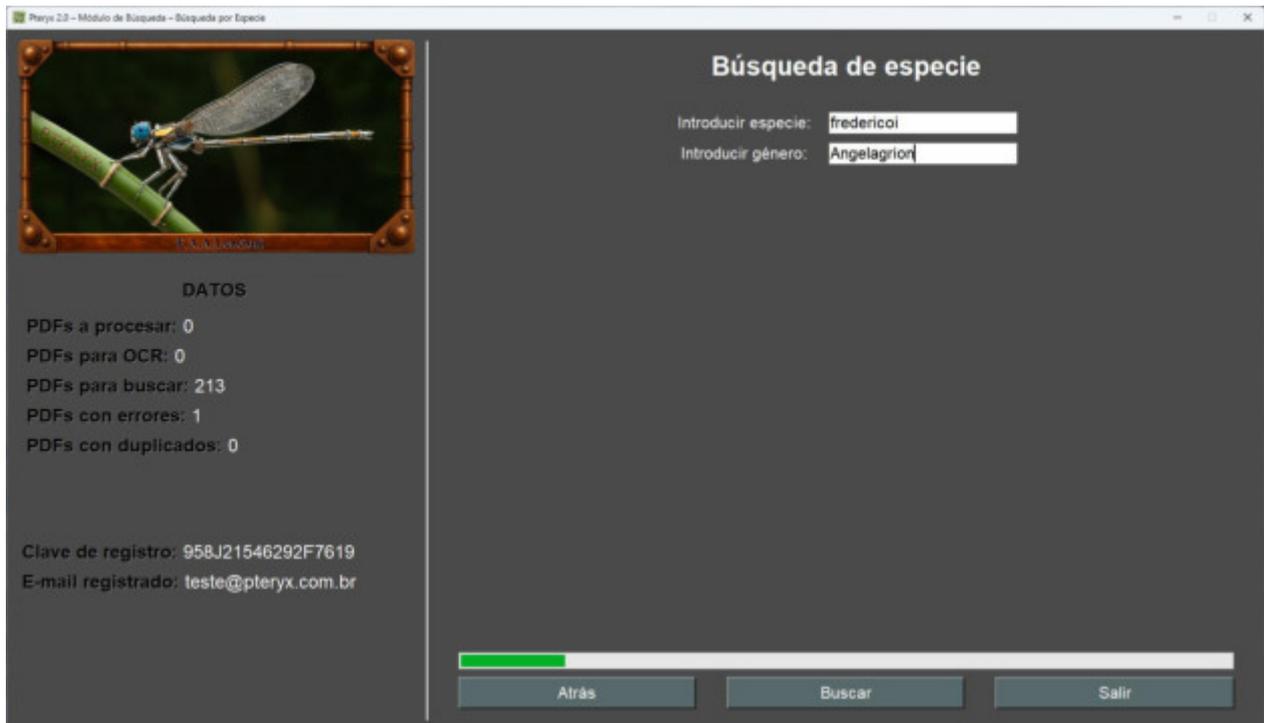


Fig. 13 – Búsqueda por especie, con género.

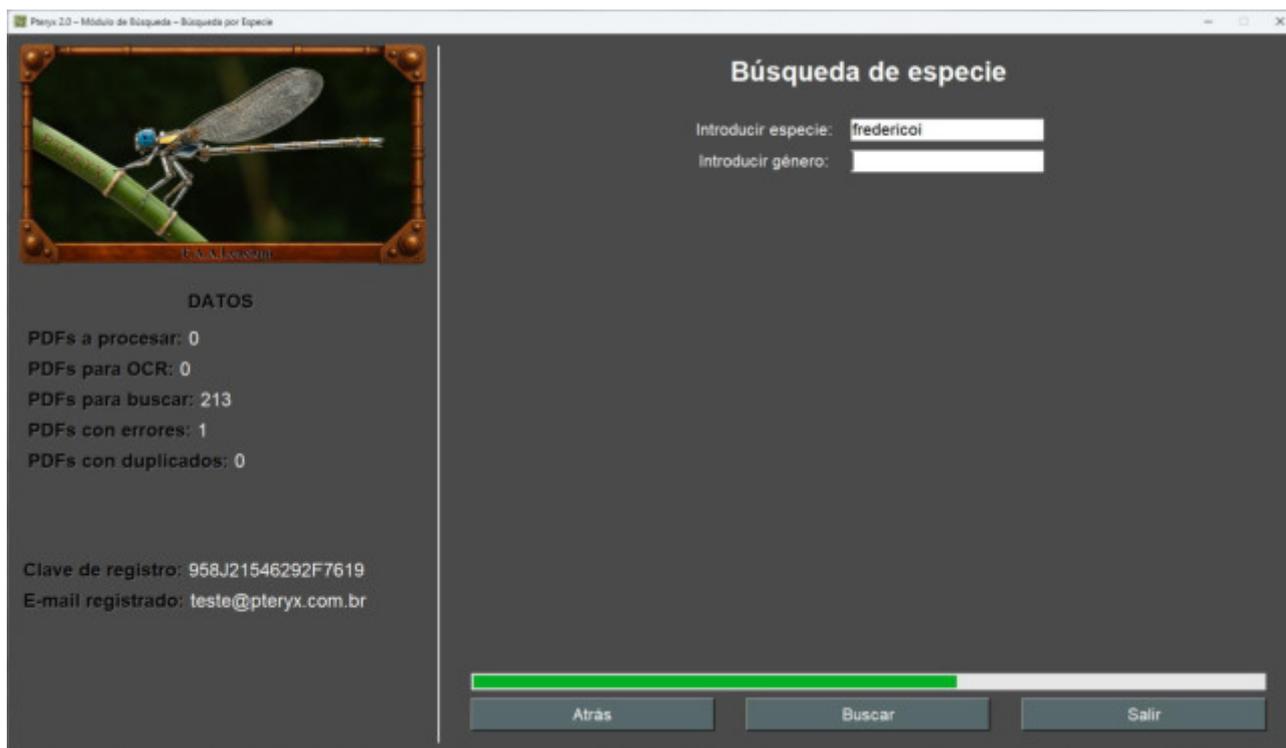


Fig. 14 – Búsqueda por especie, sin género.

5.2 Búsqueda por Lista de Especies

Esta búsqueda se basa en un **archivo Excel** (Data.xlsx – Fig. 15) que contiene múltiples especies.

En este modo de búsqueda, el usuario puede buscar de 2 a cientos de especies a la vez, lo que facilita en gran medida trabajar con grandes volúmenes de datos, como al redactar revisiones sistemáticas y tesis.

Normalmente reunimos cientos y cientos de documentos, y el análisis siempre es lento, propenso a errores, agotador y que consume mucho tiempo.

Con este modo, usted crea una simple hoja de cálculo y deja que el sistema trabaje por usted.

Si está buscando en cientos de documentos, puede dejar que el sistema trabaje mientras duerme o realiza otras actividades, y recoger los resultados cuando despierte o regrese del trabajo.

Debe crear la hoja de cálculo siguiendo los siguientes criterios:

- El archivo debe contener una **hoja** (llamada “**Search**”, preste atención a que la primera letra sea mayúscula) con las siguientes cinco columnas: **Species**, **Genera**, **Key**, **Full name** y **Name with abbreviation**.
- Hay un archivo de ejemplo incluido en el paquete (Data.xlsx) ya formateado correctamente; solo necesita completarlo con sus datos y referenciarlo cuando se le solicite.
- Los campos “**Full name**” y “**Name with abbreviation**” pueden completarse manualmente, pero si se dejan vacíos, el sistema los completará automáticamente.
- La **clave (Key)** puede ser cualquier palabra, pero no debe ser un subconjunto de otra clave (a menos que esté conectada por un guion).

Ejemplo: si se ingresan “**Coenagrionidae**” y “**Non Coenagrionidae**”, el sistema interpretará la segunda como parte de la primera. Sin embargo, si se escriben con un guion (“**Non-Coenagrionidae**”), el sistema las tratará como palabras distintas.

- Algunos nombres se utilizan rutinariamente para nombrar especies y, desafortunadamente, algunos pertenecen a géneros distintos pero tienen iniciales idénticas, como *Pantala flavescens* (una libélula), *Perca flavescens* (un pez), *Papiliotrema flavescens* (un hongo), *Parachalastinus flavescens* (un escarabajo), *Pulvinaria flavescens* (Homópteros). Todos ellos se abrevian como “**P. flavescens**” y, por lo tanto, si se encuentran en los PDFs buscados, se devolverán como positivos. Para facilitar la depuración de los datos obtenidos, hay una columna de “contexto” en la hoja de cálculo generada que facilita la identificación de estos falsos homónimos.

	A	B	C	D	E	F
	Species	Genera	Key	Full name	Name with abbreviation	
1						
2	Species_1	Genera_1	interest	Genera_1 Species_1	G. Species_1	
3	Species_2	Genera_2	interest	Genera_2 Species_2	G. Species_2	
4	Species_3	Genera_3	interest	Genera_3 Species_3	G. Species_3	
5	Species_4	Genera_4	interest	Genera_4 Species_4	G. Species_4	
6	Species_5	Genera_5	interest	Genera_5 Species_5	G. Species_5	
7	Species_6	Genera_6	interest	Genera_6 Species_6	G. Species_6	
8	Species_7	Genera_7	interest	Genera_7 Species_7	G. Species_7	
9	Species_8	Genera_8	interest	Genera_8 Species_8	G. Species_8	
10	Species_9	Genera_9	interest	Genera_9 Species_9	G. Species_9	
11	Species_10	Genera_10	interest	Genera_10 Species_10	G. Species_10	
12	Species_11	Genera_11	Proto	Genera_11 Species_11	G. Species_11	
13	Species_12	Genera_12	Proto	Genera_12 Species_12	G. Species_12	
14	Species_13	Genera_13	Proto	Genera_13 Species_13	G. Species_13	
15	Species_14	Genera_14	Proto	Genera_14 Species_14	G. Species_14	
16	Species_15	Genera_15	Proto	Genera_15 Species_15	G. Species_15	
17	Species_16	Genera_16	Proto	Genera_16 Species_16	G. Species_16	
18	Species_17	Genera_17	Proto	Genera_17 Species_17	G. Species_17	
19	Species_18	Genera_18	Proto	Genera_18 Species_18	G. Species_18	
20	Species_19	Genera_19	Proto	Genera_19 Species_19	G. Species_19	
21	Species_20	Genera_20	Proto	Genera_20 Species_20	G. Species_20	
22						
23						

Fig. 15 – Lista de especies, ejemplo de archivo.

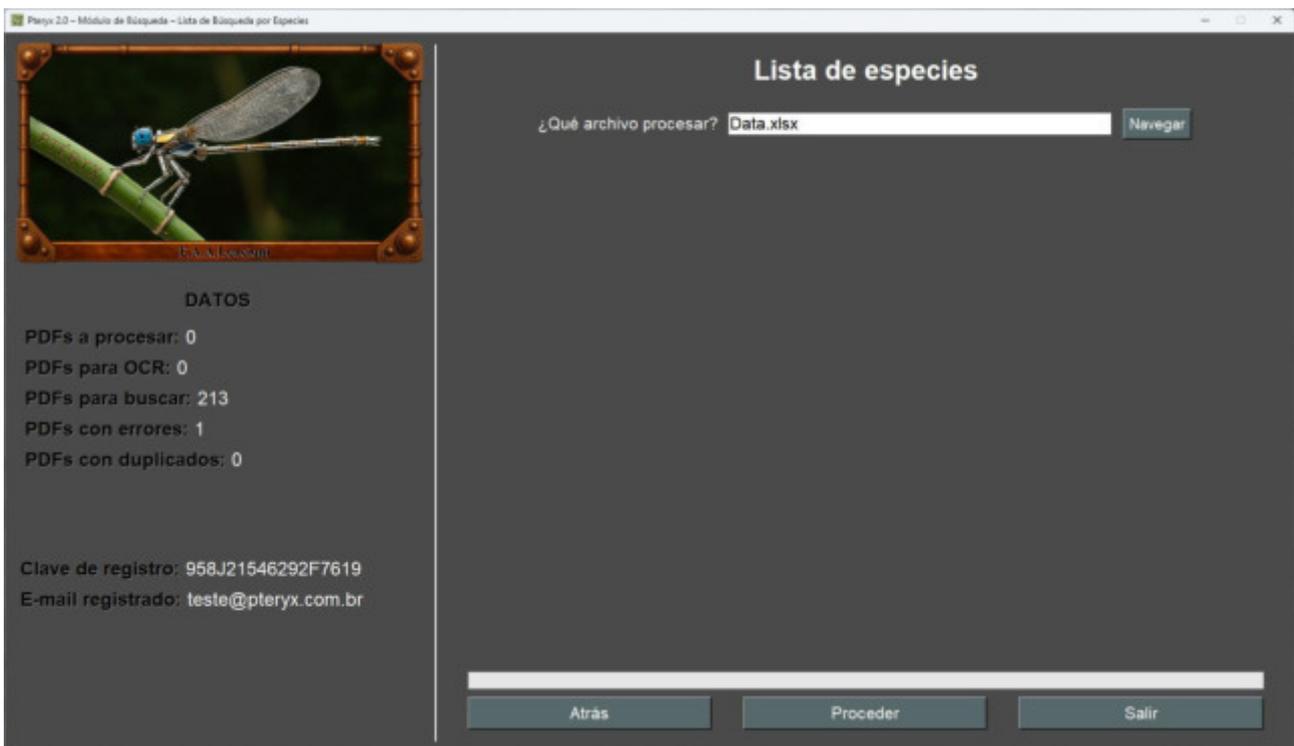


Fig. 16 – Búsqueda por lista de especies, selección del archivo.

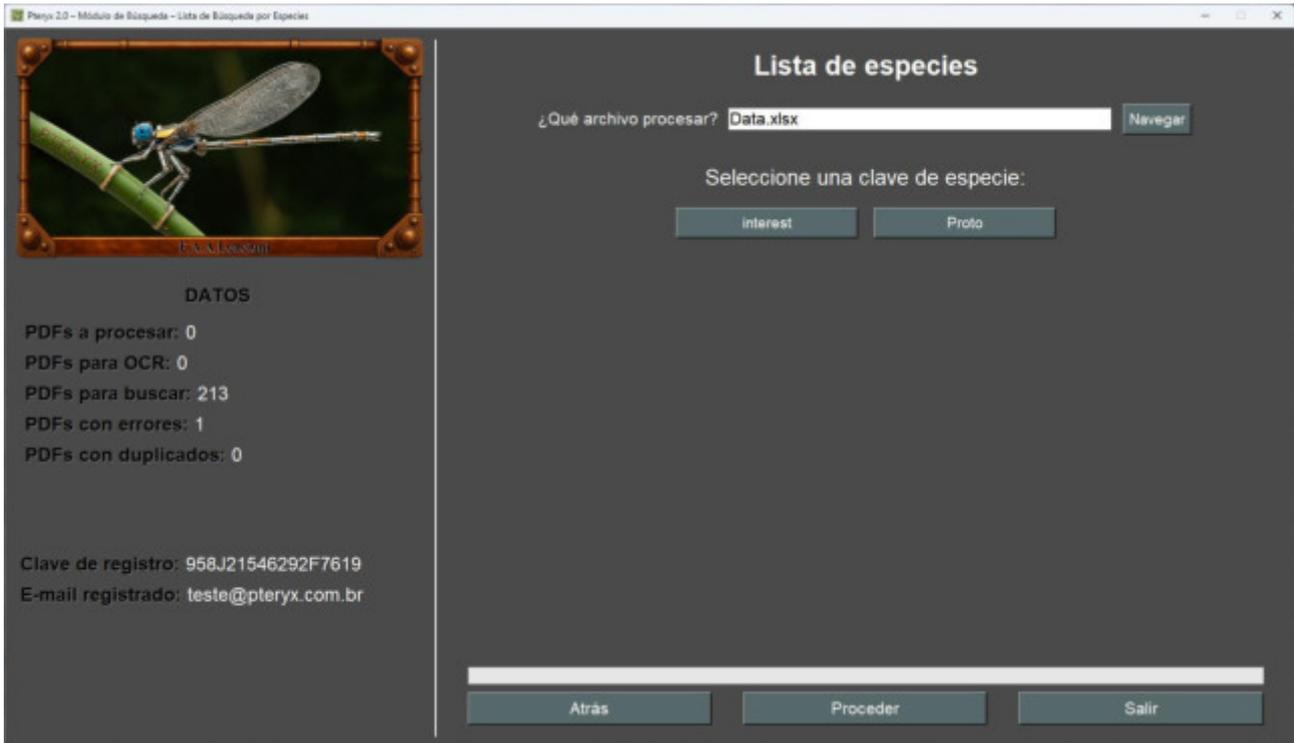


Fig. 17 – Búsqueda por lista de especies, selección de la clave a buscar; esta clave agrupa la lista de especies en el archivo **.xlsx**.

5.3 Búsqueda por Términos o Frases

Esta función localiza palabras clave o frases específicas dentro de los documentos. La búsqueda se realiza en todos los PDFs almacenados en la carpeta **3_PDFs_to_search**. El término o frase puede tener **hasta 150 caracteres** (Fig. 18).

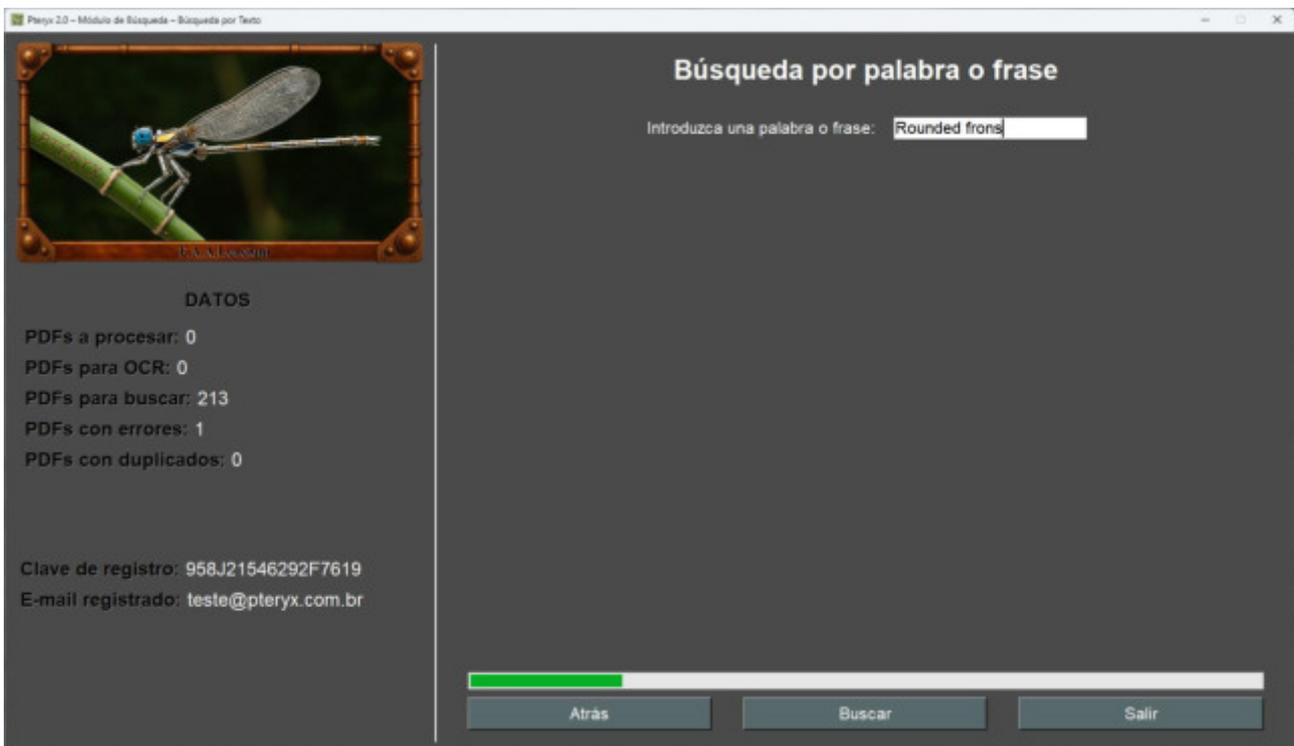


Fig. 18 – Búsqueda por palabra o frase (limitada a 150 caracteres).



7. Solución de Problemas

Error: El software no se inicia

- ✓ Asegúrese de que todos los archivos del software estén en la misma carpeta.
- ✓ Intente ejecutar el programa como Administrador.

Error: No se encontraron resultados de búsqueda

- ✓ Verifique que los PDFs hayan sido procesados correctamente en el módulo de preprocesamiento.
 - ✓ Confirme que los términos de búsqueda estén escritos correctamente.
-

8. Contacto y Soporte

Para consultas y soporte técnico, contacte por correo electrónico: sac@pteryx.com.br

9. Glosario

OCR (Reconocimiento Óptico de Caracteres)

Tecnología que convierte imágenes que contienen texto (como PDFs escaneados) en texto editable y buscable.

PDF (Formato de Documento Portátil)

Un formato de archivo utilizado para presentar documentos independientemente del software, hardware o sistema operativo.

Preprocesamiento

Etapas iniciales donde los PDFs se organizan, se revisan, se eliminan duplicados y se preparan para búsquedas o OCR.

Hash (Hashing)

Método para identificar archivos únicos generando una secuencia numérica basada en el contenido del archivo, utilizada para detectar duplicados.

Duplicados

Archivos idénticos que se identifican y separan automáticamente para evitar procesamiento redundante.

Carpeta “1_PDFs_to_Process”

Directorio donde el usuario debe colocar los PDFs originales para iniciar el procesamiento en Pteryx.

Carpeta “2_PDFs_to_OCR”

Carpeta donde se almacenan los PDFs basados en imágenes que requieren OCR.

Carpeta “3_PDFs_to_search”

Carpeta donde se almacenan los PDFs procesados y listos para ser buscados.

Búsqueda por Especie

Búsqueda de ocurrencias de una especie específica dentro de los PDFs procesados.

Búsqueda por Lista de Especies

Búsqueda simultánea de múltiples especies basada en una hoja de cálculo de Excel formateada.

Búsqueda por Términos o Frases

Búsqueda de palabras clave o expresiones dentro de los PDFs.



10. Guía Rápida — Pteryx

Paso 1: Requisitos Básicos

Antes de instalar, asegúrese de que su computadora cumpla con los requisitos mínimos:

- **Sistema Operativo:** Windows 10 o superior
 - **Procesador:** Intel Core i5 o equivalente
 - **Memoria RAM:** 8 GB o más
 - **Espacio libre en disco:** al menos 2 GB
 - **Microsoft Excel o equivalente instalado** (para la gestión de informes)
-

Paso 2: Descarga e Instalación

1. Visite el sitio web oficial de Pteryx (www.pteryx.com.br).
 2. Descargue el instalador de Pteryx.
 3. Ejecute el instalador (Pteryx_installer.exe).
 4. Durante la instalación, introduzca su correo electrónico cuando se le solicite.
 5. El instalador descargará los componentes necesarios y creará la estructura de carpetas en su computadora.
 6. Espere a que finalice la instalación.
-

Paso 3: Abrir Pteryx por Primera Vez

1. Después de la instalación, abra el programa desde el acceso directo en el escritorio o desde el menú Inicio.
 2. El software detectará automáticamente el idioma de su sistema y configurará la interfaz correspondiente.
 3. Regístrese (si se le solicita) o proceda según las instrucciones de la pantalla inicial.
-

Paso 4: Preparar sus PDFs para la Búsqueda

1. Reúna los archivos PDF que desea buscar.
 2. Copie y pegue estos archivos en la carpeta 1_PDFs_to_Process, que se creó automáticamente durante la instalación.
 - Puede encontrar esta carpeta en el directorio donde instaló Pteryx.
 3. Alternativamente, dentro del software, use la función “Reunir PDFs” para seleccionar carpetas o discos y copiar automáticamente los PDFs a la carpeta de procesamiento.
-

Paso 5: Ejecutar el Preprocesamiento

1. Desde el menú principal de Pteryx, haga clic en Preprocesamiento.
 2. Haga clic en Iniciar para comenzar a preprocesar los PDFs.
 3. El software revisará sus archivos, separándolos en carpetas específicas:
 - PDFs corruptos → 4_PDFs_with_Errors
 - Duplicados → 5_PDFs_with_duplicates
 - PDFs que requieren OCR → 2_PDFs_to_OCR
 - PDFs listos para búsqueda → 3_PDFs_to_search
 4. Si hay archivos que requieren OCR, el sistema le preguntará si desea realizar el Reconocimiento Óptico de Caracteres. Haga clic en Sí y espere a que se complete.
-



Paso 6: Realizar Búsquedas en los PDFs

1. Después del preprocesamiento, regrese al menú principal.
 2. Haga clic en Búsqueda para abrir el módulo correspondiente.
 3. Elija entre los tres tipos de búsqueda:
 - Búsqueda por Especie — para buscar una especie específica.
 - Búsqueda por Lista de Especies — para buscar múltiples especies basadas en una hoja de cálculo de Excel (Data.xlsx).
 - Búsqueda por Términos o Frases — para buscar palabras clave o frases.
 4. Configure los parámetros según el tipo de búsqueda seleccionado y haga clic en Iniciar.
 5. Espere a que se generen los informes.
-

Paso 7: Ver y Exportar Resultados

- Los informes se guardan automáticamente en la carpeta 6_PDFs_with_target_data.
 - Estarán organizados en subcarpetas según la especie o término buscado.
 - Abra los archivos .xlsx con Microsoft Excel o software compatible para un análisis detallado.
-

Consejos Finales

- Siempre realice el preprocesamiento antes de realizar búsquedas.
 - Mantenga sus carpetas organizadas para evitar errores.
 - Consulte las preguntas frecuentes para resolver problemas comunes.
 - Para soporte técnico, contacte a: sac@pteryx.com.br
-