



F.A.A.Lencioni

# Ptéryx

## Índice

1.	– <a href="#">Introdução</a> .....	1
2.	– <a href="#">Instalação e Requisitos do Sistema</a> .....	2
2.1	– <a href="#">Requisitos Mínimos</a> .....	2
2.2	– <a href="#">Instalação</a> .....	2
2.3	– <a href="#">Idiomas</a> .....	2
3	– <a href="#">Interface do Usuário</a> .....	3
3.1	– <a href="#">Pré-registro</a> .....	3
3.2	– <a href="#">Aquisição da Licença</a> .....	4
3.3	– <a href="#">Pós-registro – Tela inicial</a> .....	4
4	– <a href="#">Módulo de Pré-processamento de PDFs</a> .....	4
4.1	– <a href="#">Organização dos Arquivos</a> .....	4
4.2	– <a href="#">Processamento</a> .....	6
4.3	– <a href="#">Execução do OCR</a> .....	7
4.4	– <a href="#">Finalização</a> .....	8
4.5	– <a href="#">Prosseguir</a> .....	8
5	– <a href="#">Módulo de Busca</a> .....	9
5.1	– <a href="#">Busca por Espécie</a> .....	9
5.2	– <a href="#">Busca por Lista de Espécies</a> .....	10
5.3	– <a href="#">Busca por Termos ou Frases</a> .....	12
6	– <a href="#">Exportação e Relatórios</a> .....	13
7	– <a href="#">Solução de Problemas</a> .....	14
8	– <a href="#">Contato e Suporte</a> .....	14
9	– <a href="#">Glossário</a> .....	14
10	– <a href="#">Guia Rápido — Pteryx</a> .....	15



# 1. Introdução

## A Evolução da Pesquisa Científica e a Necessidade de Automação

Até recentemente, a pesquisa científica dependia de documentos físicos — os pesquisadores precisavam folhear páginas, anotar manualmente as informações relevantes e lidar com grandes volumes de papel. Esse processo era demorado, trabalhoso e propenso a erros humanos.

A introdução do formato PDF em 1993 revolucionou a forma como armazenamos e acessamos informações. Os documentos passaram a ser arquivados digitalmente, simplificando a organização e eliminando pilhas de papel. No entanto, mesmo com essa mudança digital, buscar manualmente em centenas ou milhares de PDFs continua sendo extremamente tedioso.

Atualmente, a maioria dos artigos e livros acadêmicos está disponível online, reduzindo drasticamente a circulação de materiais físicos. No Brasil, o impacto é evidente: na última década, 17% das bancas de jornal e 29% das livrarias fecharam, e entre 2015 e 2020, pelo menos 764 bibliotecas públicas foram encerradas (Sistema Nacional de Bibliotecas Públicas). Isso força os pesquisadores a dependerem de dispositivos eletrônicos para leitura — o tempo prolongado diante das telas pode causar fadiga visual e outros problemas oculares.

## Os Desafios da Busca Manual em PDFs

Buscar manualmente documentos digitais envolve abrir arquivos, ler cada página, anotar informações e repetir continuamente esse processo. Esse método, além de extremamente demorado, aumenta a probabilidade de erros, já que a fadiga visual pode levar à omissão de dados relevantes. Estudos indicam que a leitura em telas retro iluminadas exige maior esforço visual, impactando na precisão das análises e tornando o processo menos eficiente.

Além da sobrecarga cognitiva, a repetição contínua dessas tarefas exaure física e mentalmente o pesquisador. Considerando que a leitura de um PDF de 5 páginas leva em média 5 minutos, um profissional precisaria de 8 horas para revisar aproximadamente 100 documentos. Para analisar 3.000 arquivos, seriam necessários cerca de 30 dias de trabalho ininterrupto. Por que realizar essa tarefa manualmente quando um software pode executá-la em poucos minutos, com maior precisão e sem interrupções?

## Limitações da Busca do Windows e a Solução Pteryx

O Microsoft Windows permite buscar informações dentro de múltiplos arquivos, mas essa funcionalidade possui limitações: permite buscas por apenas um termo de cada vez, não gera relatórios detalhados e não organiza automaticamente os arquivos. Isso torna a triagem manual lenta e ineficaz para pesquisas científicas.

O Pteryx foi desenvolvido para resolver esse problema. Criado por um pesquisador em zoologia, oferece uma solução robusta e automatizada para analisar documentos científicos em formato PDF. O software pré-processa arquivos, realiza buscas automatizadas e gera relatórios detalhados, indicando onde cada informação foi encontrada e em qual contexto. Isso torna a triagem documental mais rápida, precisa e eficiente, permitindo que pesquisadores e acadêmicos foquem na análise de dados em vez de desperdiçar tempo com tarefas repetitivas.

Automatizar essa etapa é essencial para acompanhar a crescente digitalização da ciência e garantir que a pesquisa continue avançando sem os obstáculos do trabalho manual.



## 2. Instalação e Requisitos do Sistema

### 2.1 Requisitos Mínimos

- **Sistema Operacional:** Windows 10 ou superior
- **Processador:** Intel Core i5 ou equivalente
- **Memória RAM:** 8 GB ou superior
- **Espaço em Disco:** 2 GB livres para instalação
- **Resolução mínima:** 1280 x 720
- **Software Requerido:** Microsoft Excel ou equivalente instalado (para manipulação de relatórios)
- **Bibliotecas Requeridas:** O software inclui todas as dependências no executável, não sendo necessária instalação manual.

### 2.2 Instalação

1. Baixe o instalador do Pteryx no site oficial.
2. Execute o **Pteryx\_installer.exe**, escolha a unidade de instalação e insira seu e-mail. O instalador fará o download do aplicativo e criará a estrutura de pastas necessária.

### 2.3 Idiomas

O Pteryx detecta automaticamente o idioma do seu sistema e funcionará em um dos seguintes: Inglês, Português, Espanhol, Galego, Búlgaro, Francês, Russo, Alemão, Japonês ou Mandarim. Para qualquer outro idioma, a interface padrão será o inglês.

---



### 3. Interface do Usuário

#### 3.1 Pré-registro

Antes de comprar e registrar o programa, ele funcionará no modo demonstração, exibindo um botão "registre-se agora" na tela principal (Fig. 1). No modo demo, o pré-processamento não funciona e as buscas são limitadas a três nas modalidades "Busca por espécie" e "Palavra ou frase".

Após concluir as três buscas, o programa irá diretamente para a tela de registro (Fig. 2).

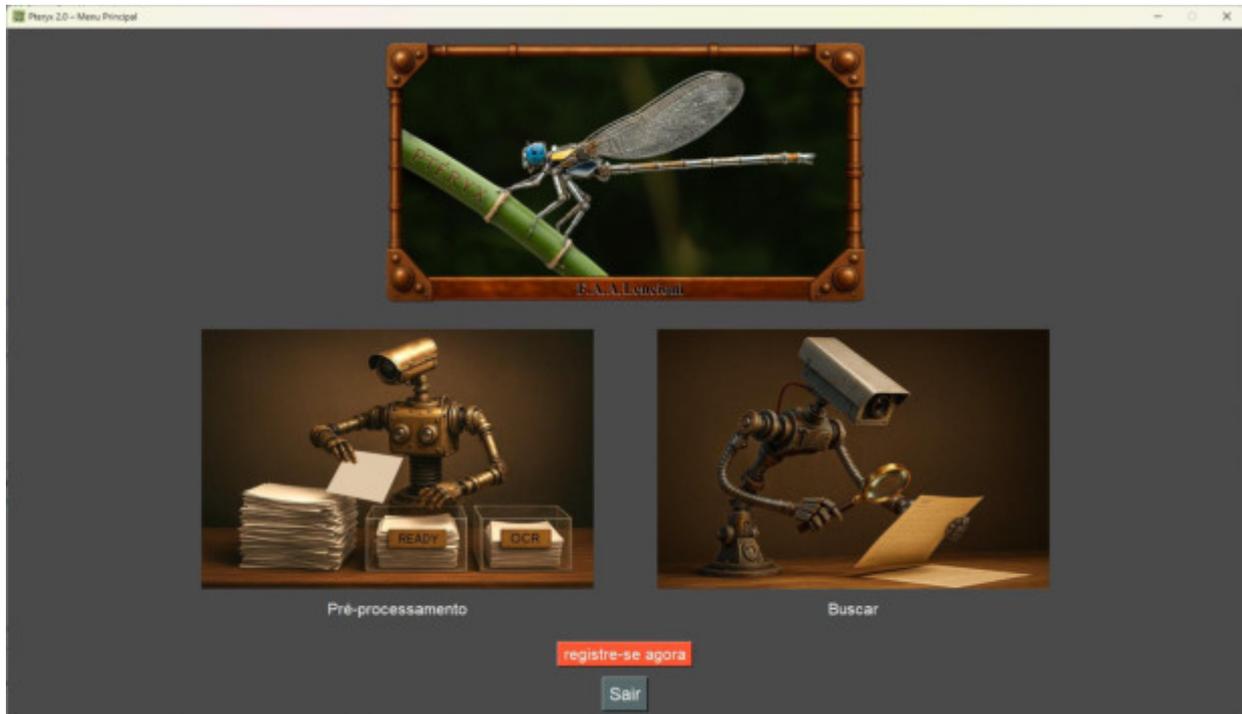


Fig. 1 – Menu Principal (tela de pré-registro).

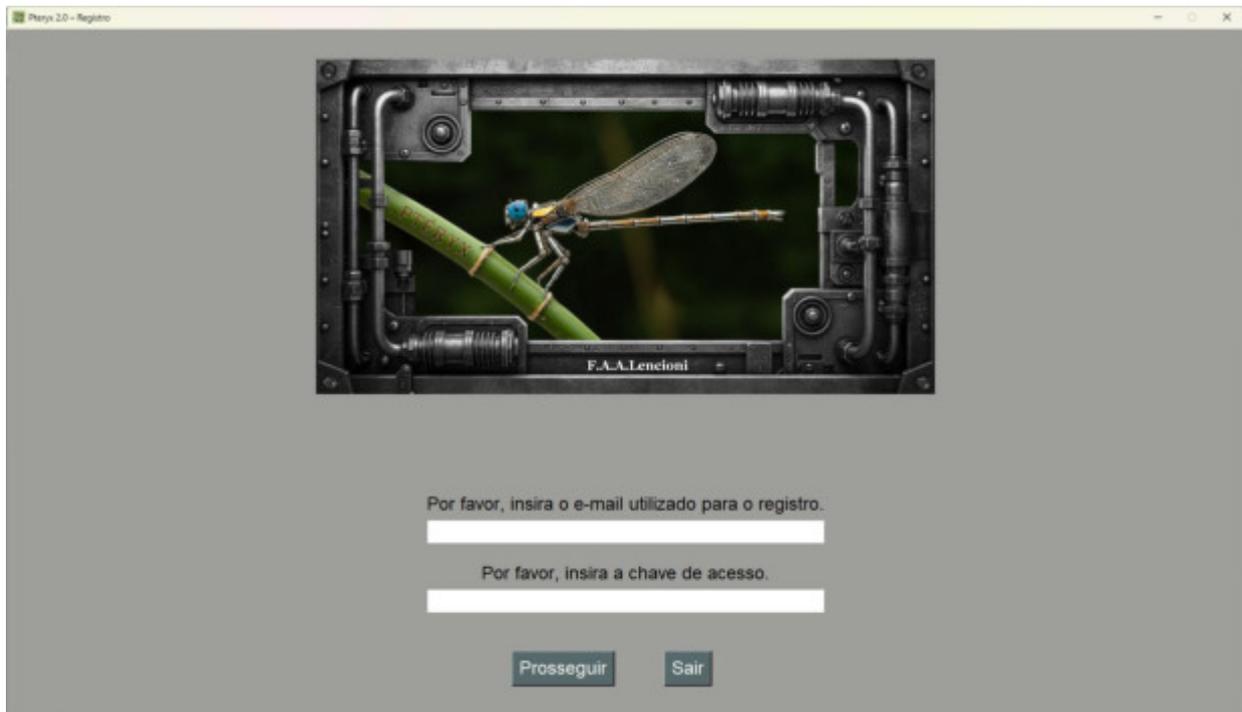


Fig. 2 – Tela de Registro.



### 3.2 Aquisição da Licença

Para adquirir a licença, acesse o site [www.pteryx.com.br](http://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 Pós-registro – Tela Inicial

Após o registro, você acessará a tela principal (Fig. 3). A partir daqui, pode selecionar um dos dois módulos:

- **Pré-processamento** (Pré-processamento de PDFs)
- **Busca (Módulo de Busca)**

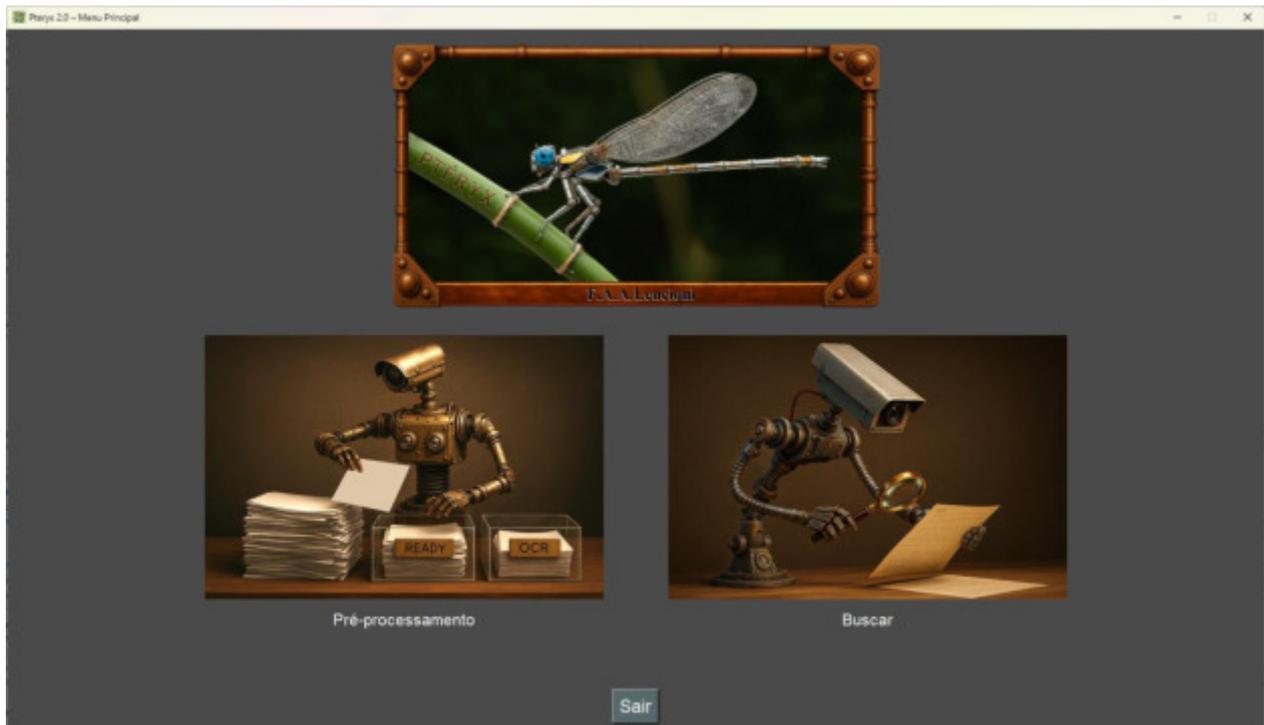


Fig. 3 – Menu Principal.

## 4. Módulo de Pré-processamento de PDFs

O software é projetado para buscar PDFs e gerar relatórios detalhados dessas buscas. Para isso, os PDFs precisam ser legíveis e pesquisáveis, livres de erros e organizados em um local predeterminado.

O sistema de pré-processamento realiza isso em duas etapas:

**Etapa 1:** Busca e copia todos os PDFs de uma pasta ou disco para o local onde serão processados (o usuário pode optar por colocar manualmente os PDFs na pasta **1\_PDFs\_to\_Process**).

**Etapa 2:** Faz a triagem de todos os PDFs, procurando erros, duplicatas e PDFs baseados em imagem (que são movidos para a pasta **2\_PDFs\_to\_OCR**).

### 4.1 Organização dos Arquivos

- Clique no botão "**Pré-processamento**" na tela principal.
- Você pode copiar e colar manualmente os PDFs na pasta **1\_PDFs\_to\_Process** ou clicar em "**Coletar PDFs**" e especificar quais pastas ou discos o programa deve coletar os PDFs.
  - O software apenas copiará seus arquivos; a estrutura original será mantida intacta.
  - Se você escolher múltiplas pastas que contêm cópias do mesmo PDF, elas serão salvas com o sufixo (1). Se houver mais de uma cópia, serão salvas com números sequenciais (2, 3, 4, etc.).

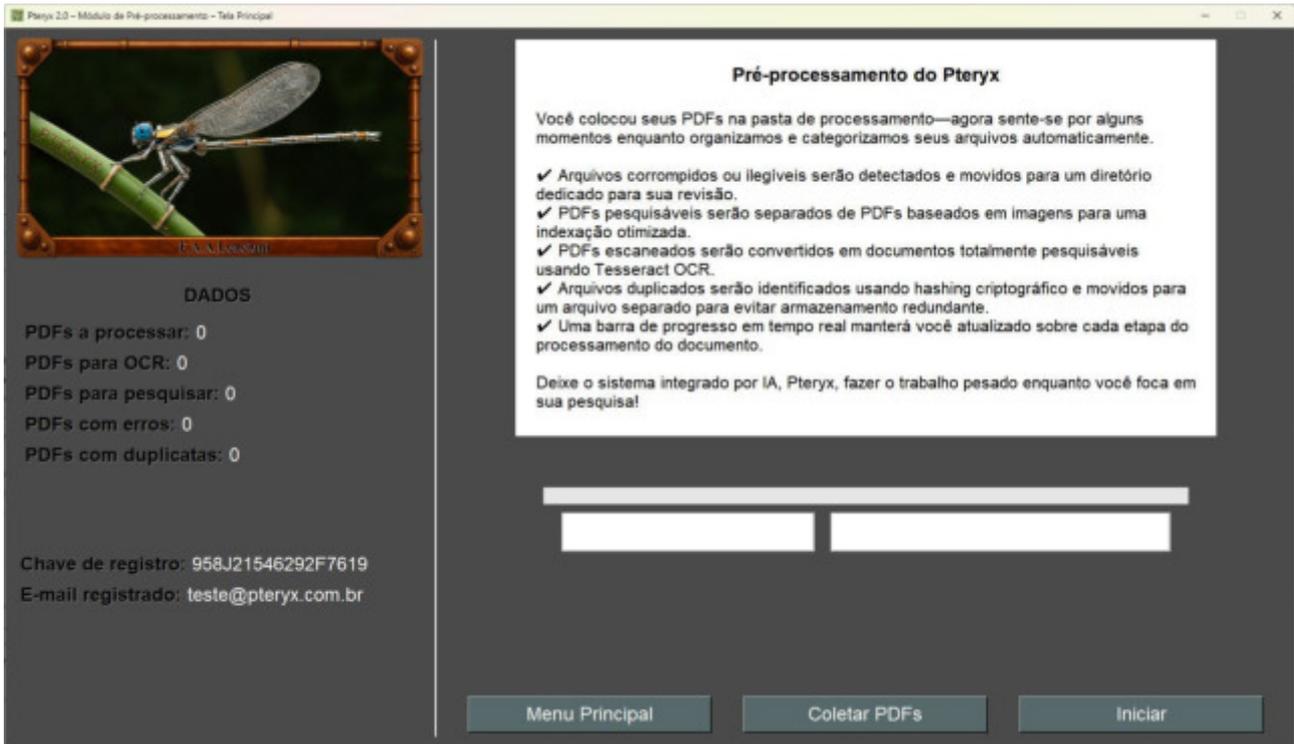


Fig. 4 – Pré-processamento de PDFs.

- Clique no botão "Coletar PDFs".

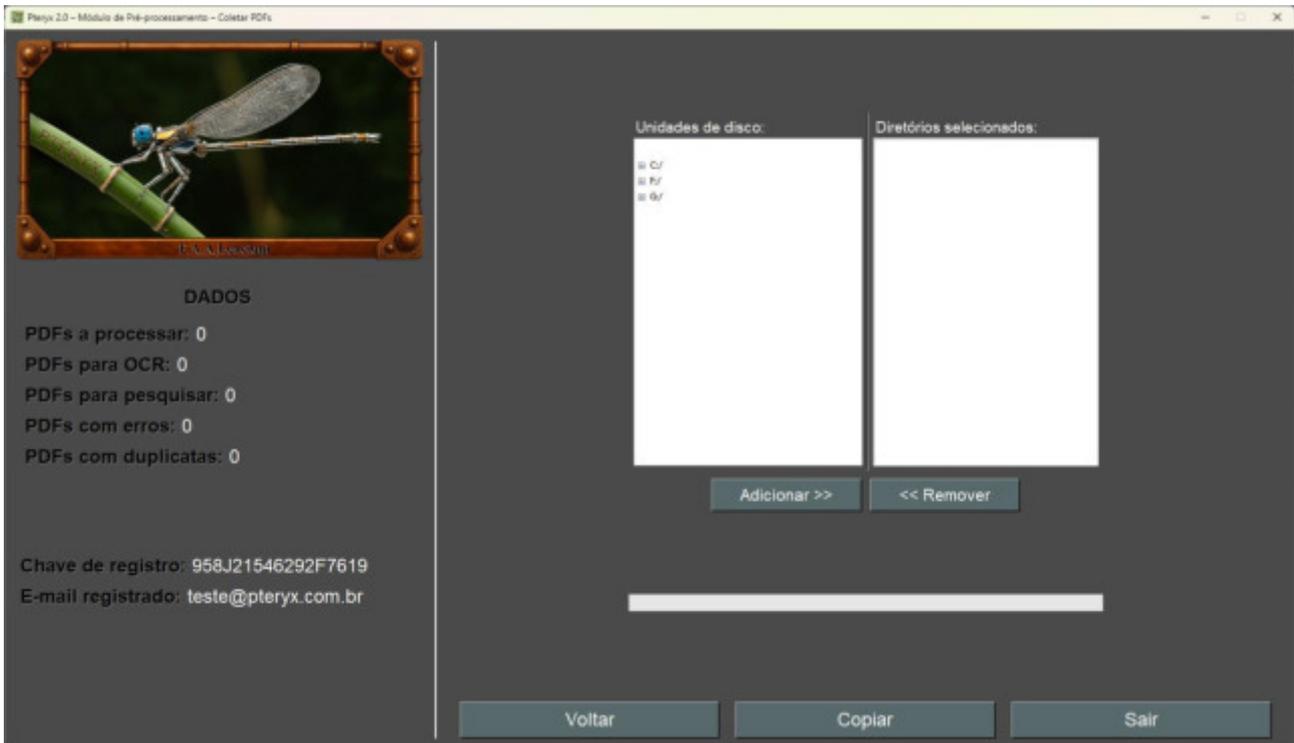


Fig. 5 – Coletar PDFs, escolhendo pastas.

- Escolha as pastas ou discos de onde os PDFs serão coletados e adicione-os aos “**Diretórios selecionados**”.
- Pressione Copiar e todos os arquivos PDF das pastas ou discos selecionados serão copiados para **1\_PDFs\_to\_Process**.

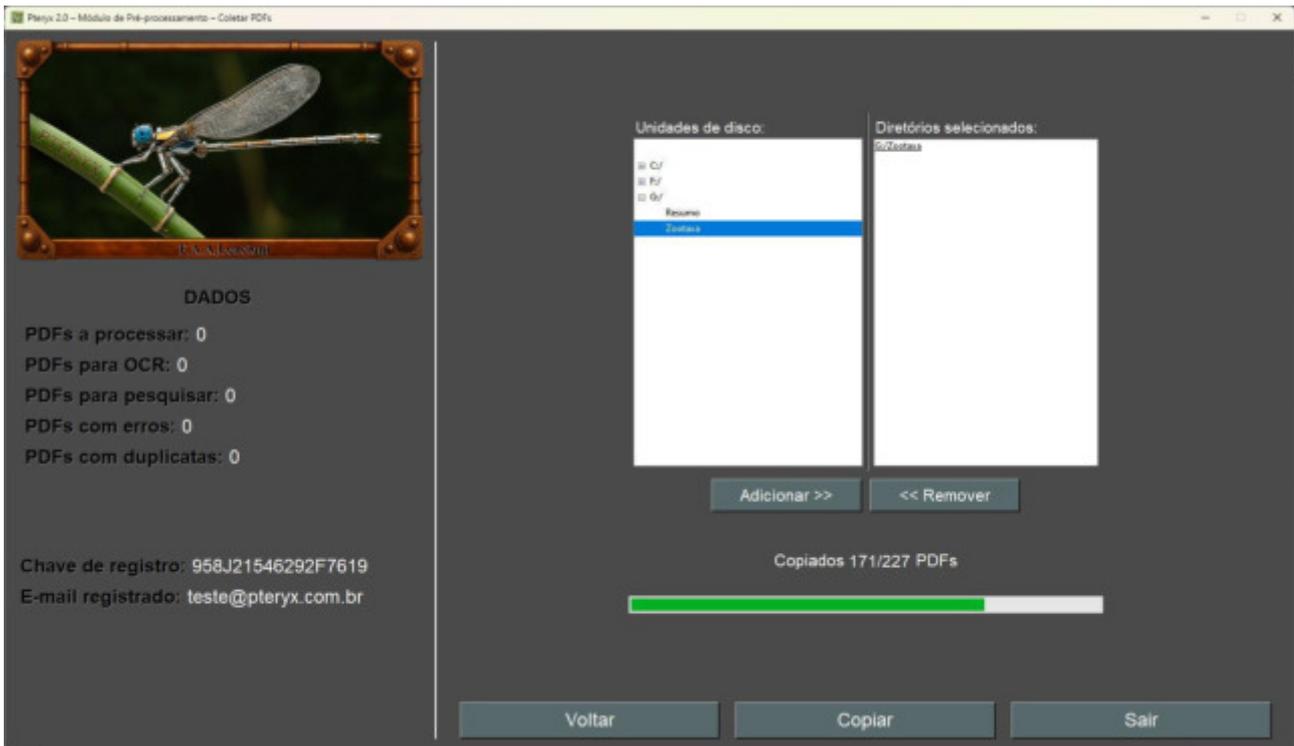


Fig. 6 – Coletar PDFs, copiando arquivos.

## 4.2 Processamento

Depois que todos os PDFs estiverem na pasta **1\_PDFs\_to\_Process**, volte ao menu principal do pré-processamento e clique em "**Iniciar**". O sistema processará todos os arquivos, verificando:

- Arquivos corrompidos ou com erro → movidos para **4\_PDFs\_with\_Errors** (Fig. 7)
- Cópias de arquivos duplicados → movidas para **5\_PDFs\_with\_duplicates** (Fig. 8)

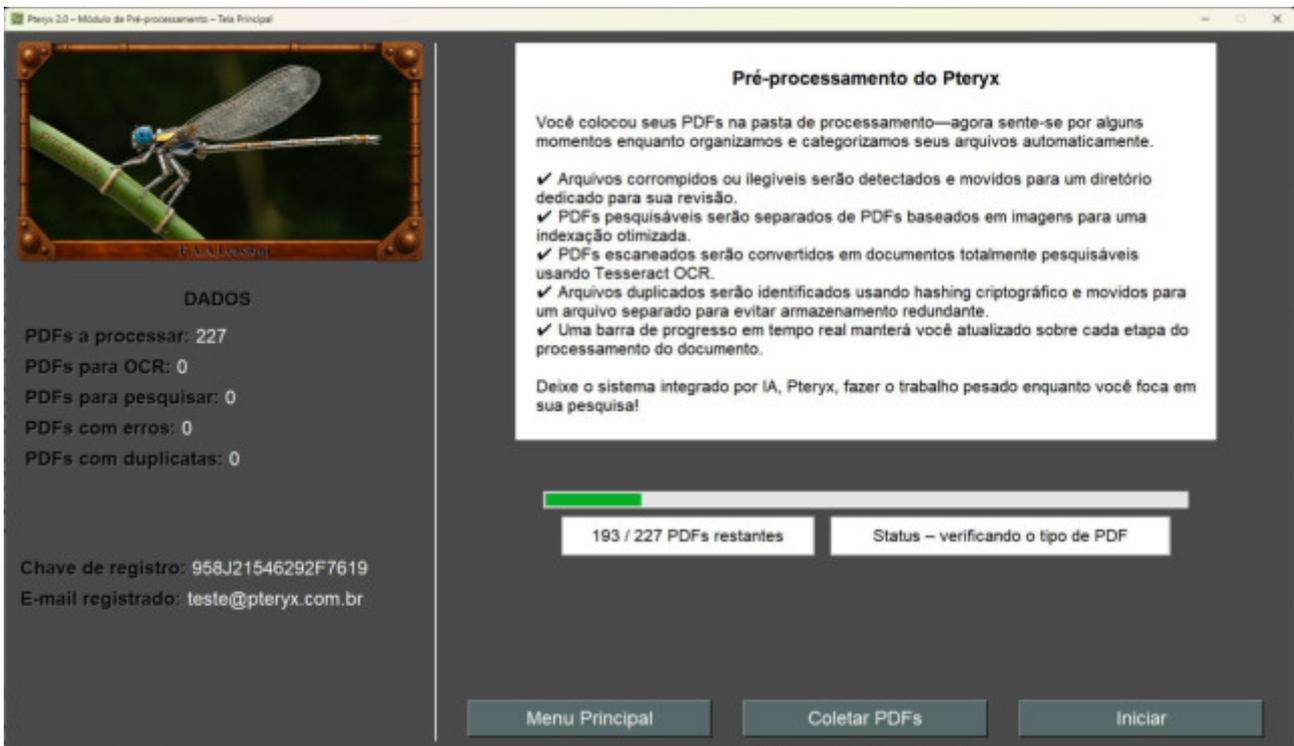


Fig. 7 – Verificando tipo de arquivo.

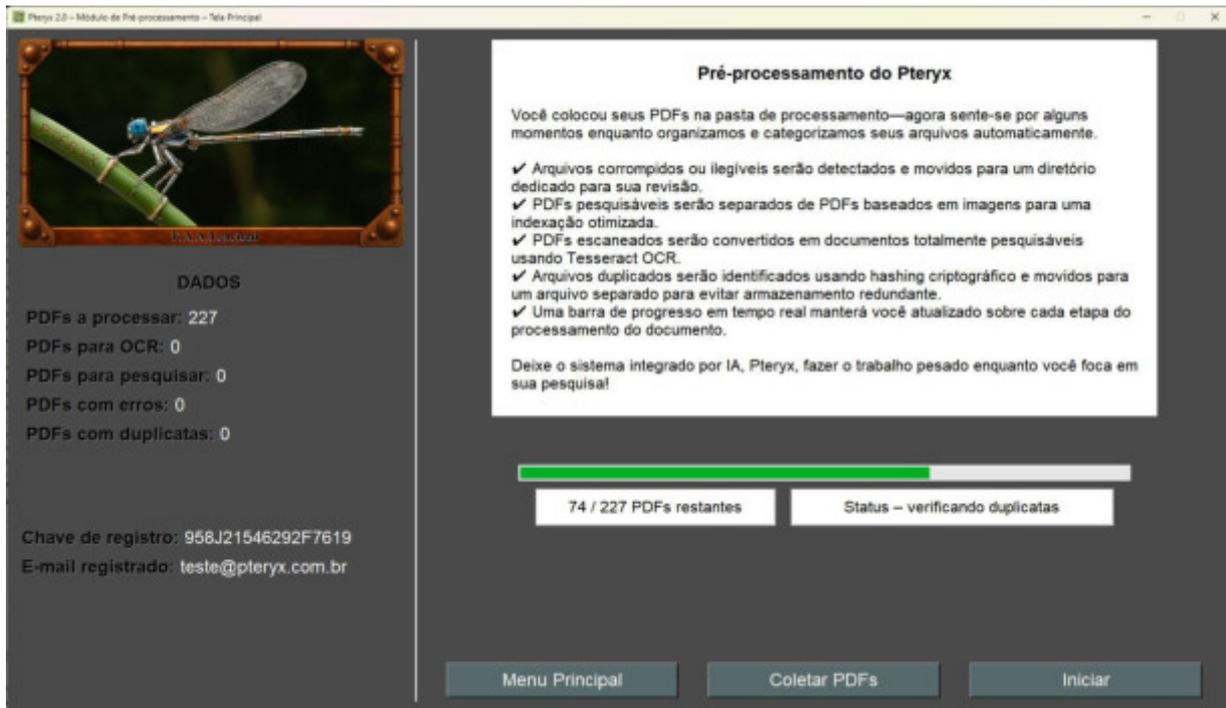


Fig. 8 – Verificando arquivos duplicados.

- PDFs baseados em imagem que requerem OCR → movidos para **2\_PDFs\_to\_OCR**
- PDFs prontos para busca → movidos para **3\_PDFs\_to\_search**

### 4.3 Execução do OCR

Depois que o pré-processamento for concluído, se houver arquivos que exigem OCR (PDFs baseados em imagem), o sistema mostrará a quantidade de arquivos e perguntará se deseja executar o OCR (Fig. 9). Se escolher **Sim**, aguarde o processamento de todos os arquivos (Fig. 10).

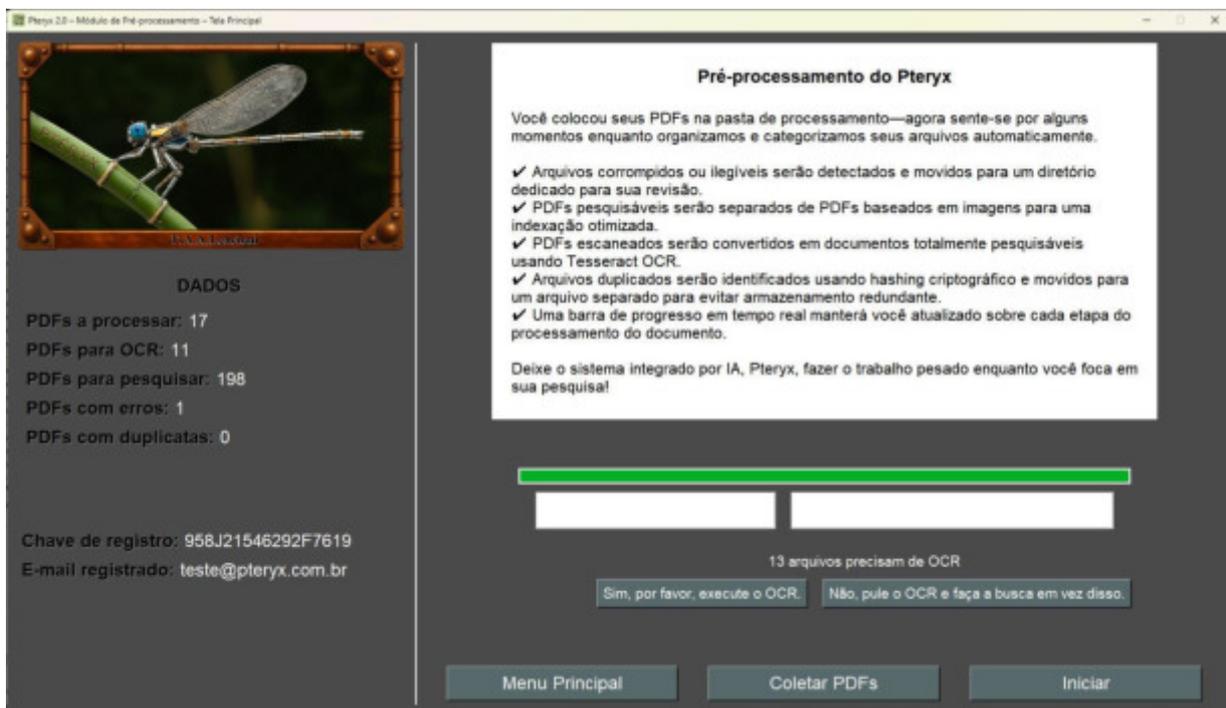


Fig. 9 – Prompt de OCR.

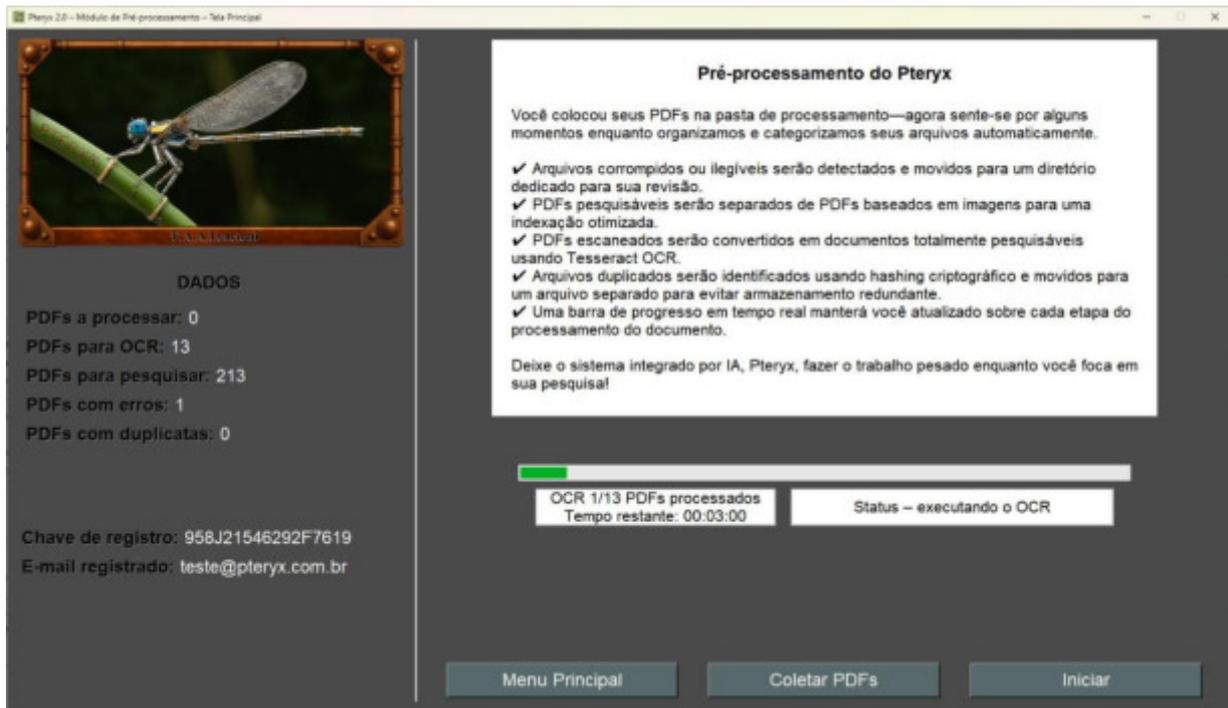


Fig. 10 – Executando OCR.

#### 4.4 Finalização

Após a conclusão do processo de OCR, os arquivos serão movidos para a pasta **3\_PDFs\_to\_search** e um relatório de todos os PDFs processados será exibido (Fig. 11).

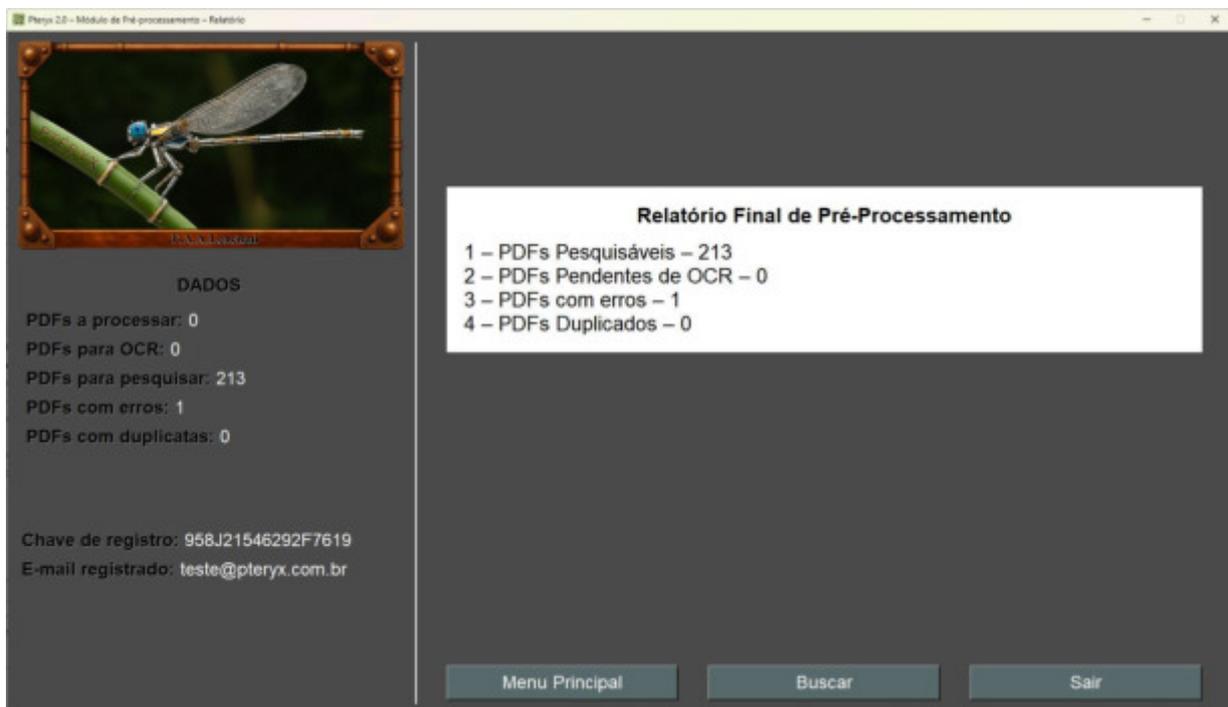


Fig. 11 – Relatório Final do Pré-processamento.

#### 4.5 Prosseguir

Agora, volte ao Menu Principal para realizar suas buscas.



## 5. Módulo de Busca

Este é o módulo principal, oferecendo três métodos de busca. Cada um gera relatórios detalhados, fáceis de revisar (Fig. 12).

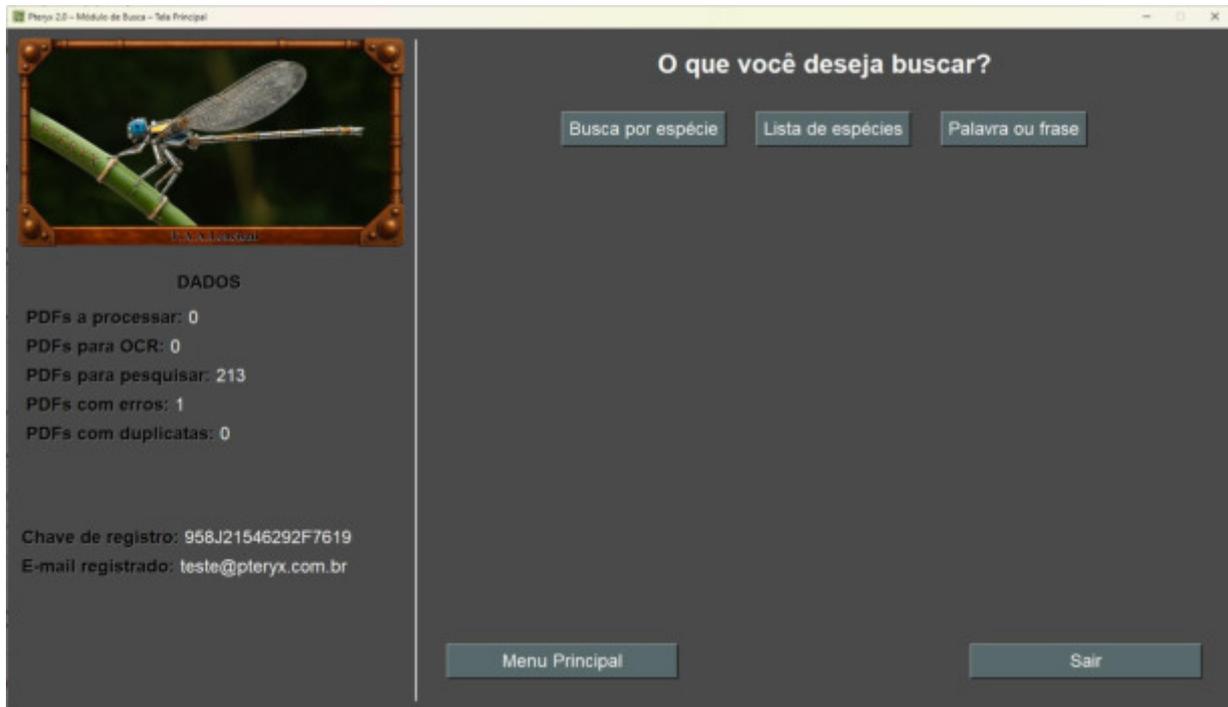


Fig. 12 – Módulo de busca.

### 5.1 Busca por Espécie

Esta função busca em todos os PDFs armazenados na pasta **3\_PDFs\_to\_search** ocorrências de uma espécie específica. A busca pode ser realizada com a entrada do gênero (Fig. 13) ou sem ela (Fig. 14).

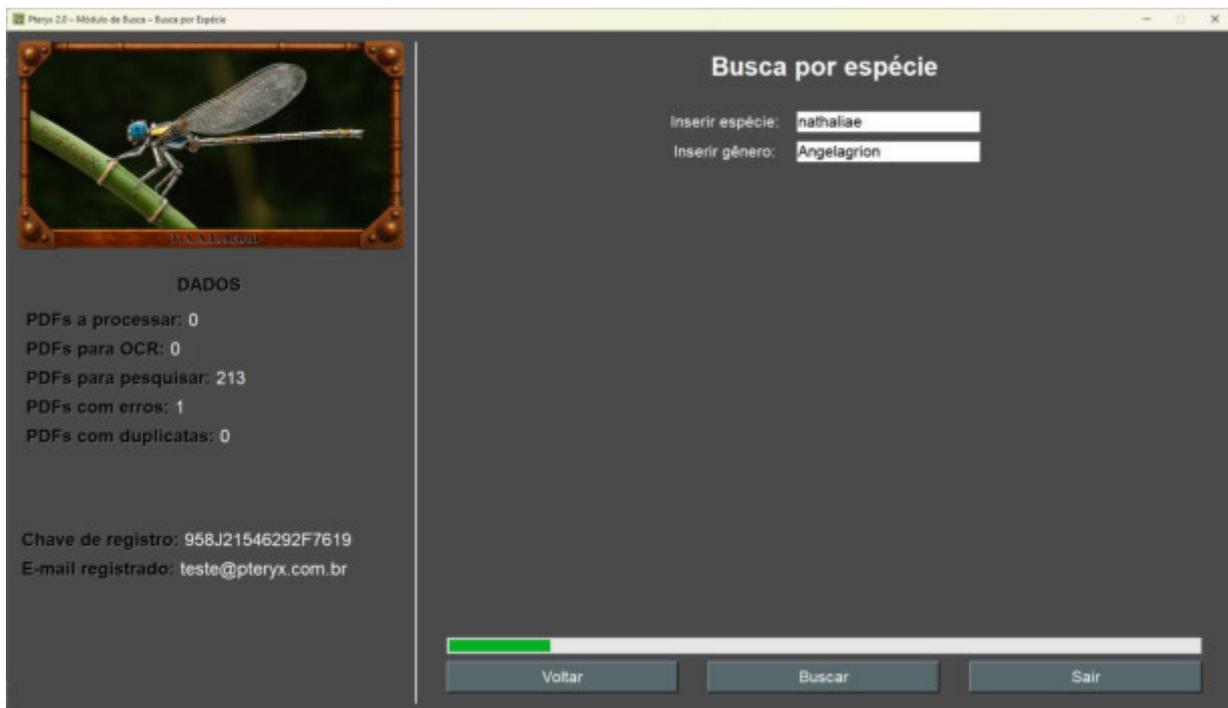


Fig. 13 – Busca por espécie, com gênero.

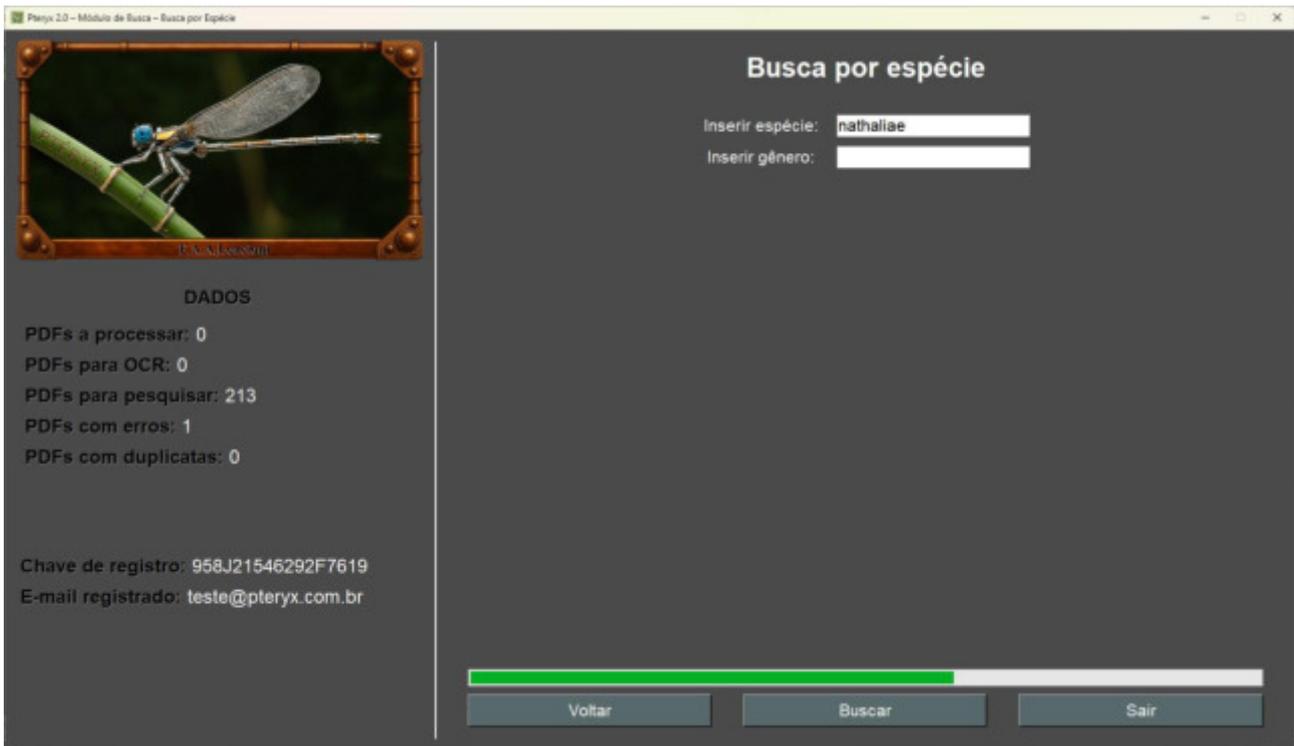


Fig. 14 – Busca por espécie, sem gênero.

## 5.2 Busca por Lista de Espécies

Esta busca é baseada em um **arquivo Excel** (Data.xlsx – Fig. 15) contendo múltiplas espécies.

Neste modo de busca, o usuário pode procurar de 2 até centenas de espécies de uma só vez, o que facilita muito o trabalho com grandes volumes de dados, como ao escrever revisões sistemáticas e teses. Normalmente reunimos centenas e centenas de documentos, e a análise sempre é lenta, sujeita a erros, cansativa e demorada.

Com este modo, você cria uma planilha simples e deixa o sistema trabalhar por você. Se estiver pesquisando em centenas de documentos, pode deixar o sistema funcionando enquanto dorme ou faz outras atividades, e colher os resultados ao acordar ou retornar do trabalho.

Você precisa criar a planilha seguindo os seguintes critérios:

- O arquivo deve conter uma **aba** (nomeada "**Search**" – atenção: com inicial maiúscula) com as seguintes cinco colunas: **Species**, **Genera**, **Key**, **Full name** e **Name with abbreviation**.
- Há um arquivo de exemplo no pacote (**Data.xlsx**) já formatado corretamente; você só precisa preencher os dados e referenciá-lo quando solicitado.
- Os campos "**Full name**" (nome completo) e "**Name with abbreviation**" (nome com abreviação) podem ser preenchidos manualmente, mas se deixados em branco, o sistema os completará automaticamente.
- A **chave (Key)** pode ser qualquer palavra, mas não deve ser um subconjunto de outra chave (a menos que estejam conectadas por hífen). Exemplo: se forem inseridas **Coenagrionidae** e **Non Coenagrionidae**, o sistema interpretará a segunda como parte da primeira. Mas se forem escritas com hífen (**Non-Coenagrionidae**), o sistema as tratará como palavras distintas.
- Alguns nomes são rotineiramente usados para nomear espécies, e infelizmente algumas pertencem a gêneros distintos mas possuem iniciais idênticas, como: *Pantala flavescens* (uma libélula), *Perca flavescens* (um peixe), *Papiliotrema flavescens* (um fungo), *Parachalastinus flavescens* (um besouro), *Pulvinaria flavescens* (Homoptera), etc. Todos serão abreviados como "**P. flavescens**" e, portanto, se forem encontrados nos PDFs pesquisados, retornarão como positivos. Para facilitar a triagem dos dados obtidos, há uma coluna "contexto" na planilha gerada que facilita a identificação desses falsos homônimos.



	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 espécies, exemplo de arquivo.

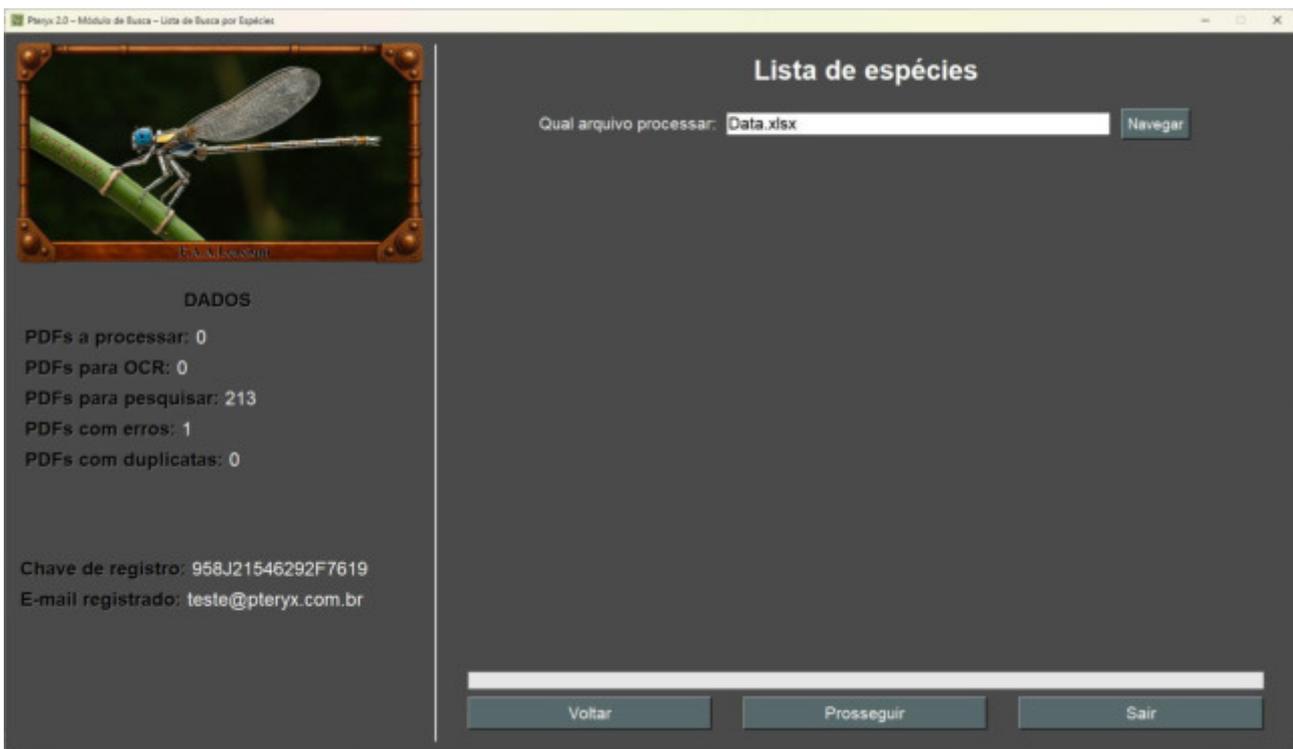
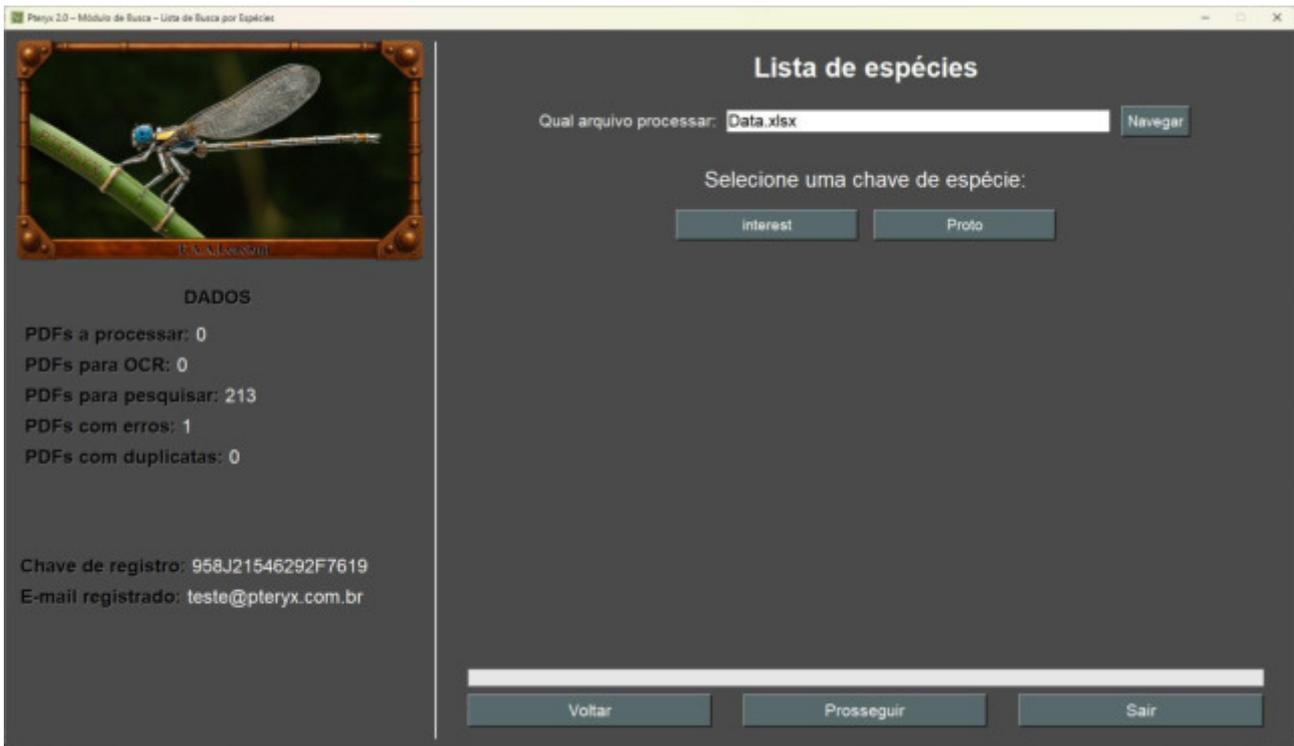


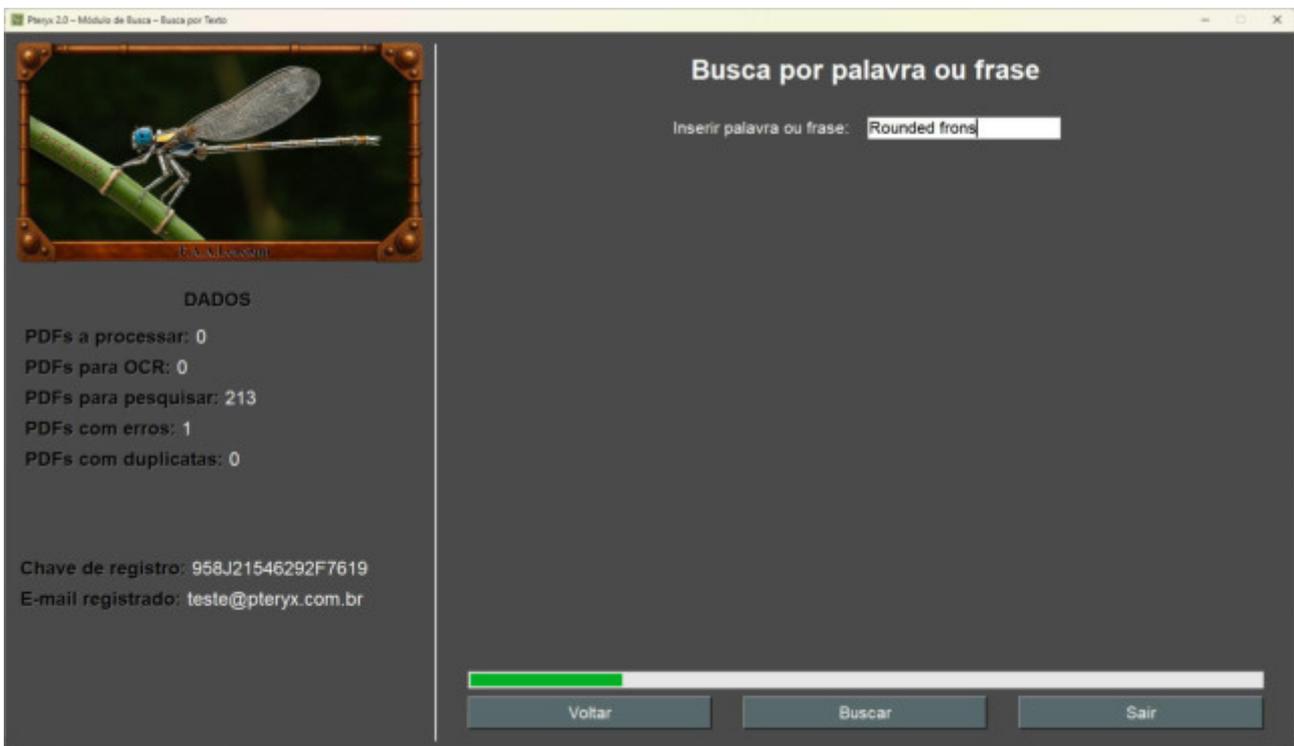
Fig. 16 – Busca por lista de espécies, selecione o arquivo.



**Fig. 17** – Busca por lista de espécies, selecione a chave a ser buscada; essa chave agrupa a lista de espécies no arquivo .xlsx.

### 5.3 Busca por Termos ou Frases

Esta função localiza **palavras-chave ou frases** específicas dentro dos documentos. A busca é realizada em todos os PDFs armazenados na pasta **3\_PDFs\_to\_search**. O termo ou frase pode ter **até 150 caracteres** (Fig. 18).



**Fig. 18** – Busca por palavra ou frase (limitada a 150 caracteres).





## 7. Solução de Problemas

**Erro:** O software não inicia

- ✓ Certifique-se de que todos os arquivos do software estejam na mesma pasta.
- ✓ Tente executar o programa como Administrador.

**Erro:** Nenhum resultado de busca encontrado

- ✓ Verifique se os PDFs foram corretamente processados no módulo de Pré-processamento.
  - ✓ Confirme se os termos de busca foram digitados corretamente.
- 

## 8. Contato e Suporte

Para dúvidas e suporte técnico, entre em contato via e-mail: [sac@pteryx.com.br](mailto:sac@pteryx.com.br)

---

## 9. Glossário

### OCR (Reconhecimento Óptico de Caracteres)

Tecnologia que converte imagens contendo texto (como PDFs digitalizados) em texto pesquisável e editável.

### PDF (Formato de Documento Portátil)

Formato de arquivo usado para apresentar documentos independentemente de software, hardware ou sistema operacional.

### Pré-processamento

Etapa inicial onde os PDFs são organizados, verificados, duplicatas removidas e preparados para buscas ou OCR.

### Hash (Hashing)

Método para identificar arquivos únicos gerando uma sequência numérica com base no conteúdo do arquivo, usada para detectar duplicatas.

### Duplicatas

Arquivos idênticos que são automaticamente identificados e separados para evitar processamento redundante.

### Pasta “1\_PDFs\_to\_Process”

Diretório onde o usuário deve colocar os PDFs originais para iniciar o processamento no Pteryx.

### Pasta “2\_PDFs\_to\_OCR”

Pasta onde são armazenados os PDFs baseados em imagem que requerem OCR.

### Pasta “3\_PDFs\_to\_search”

Pasta onde são armazenados os PDFs processados e prontos para busca.

### Busca por Espécie

Busca por ocorrências de uma espécie específica dentro dos PDFs processados.

### Busca por Lista de Espécies

Busca simultânea por múltiplas espécies com base em uma planilha Excel formatada.

### Busca por Termos ou Frases

Busca por palavras-chave ou expressões dentro dos PDFs.



## 10. Guia Rápido — Pteryx

### Etapa 1: Requisitos Básicos

Antes de instalar, verifique se seu computador atende aos requisitos mínimos:

- **Sistema Operacional:** Windows 10 ou superior
  - **Processador:** Intel Core i5 ou equivalente
  - **Memória RAM:** 8 GB ou mais
  - **Espaço livre em disco:** pelo menos 2 GB
  - **Microsoft Excel ou equivalente instalado** (para manipulação de relatórios)
- 

### Etapa 2: Download e Instalação

1. Acesse o site oficial do Pteryx ([www.pteryx.com.br](http://www.pteryx.com.br)).
  2. Baixe o instalador do Pteryx.
  3. Execute o instalador (Pteryx\_installer.exe).
  4. Durante a instalação, insira seu e-mail quando solicitado.
  5. O instalador fará o download dos componentes necessários e criará a estrutura de pastas em seu computador.
  6. Aguarde a conclusão da instalação.
- 

### Etapa 3: Abrindo o Pteryx pela Primeira Vez

1. Após a instalação, abra o programa pelo atalho na área de trabalho ou no menu Iniciar.
  2. O software detectará automaticamente o idioma do seu sistema e configurará a interface correspondente.
  3. Registre-se (se solicitado) ou prossiga conforme as instruções da tela inicial.
- 

### Etapa 4: Preparando seus PDFs para Busca

1. Reúna os arquivos PDF que deseja pesquisar.
  2. Copie e cole esses arquivos na pasta 1\_PDFs\_to\_Process, que foi criada automaticamente durante a instalação.
    - Você pode encontrar esta pasta no diretório onde instalou o Pteryx.
  3. Alternativamente, dentro do software, utilize a função “Reunir PDFs” para selecionar pastas ou discos e copiar automaticamente os PDFs para a pasta de processamento.
- 

### Etapa 5: Executando o Pré-processamento

1. No menu principal do Pteryx, clique em Pré-processamento.
  2. Clique em Iniciar para começar o pré-processamento dos PDFs.
  3. O software verificará seus arquivos, separando-os em pastas específicas:
    - PDFs corrompidos → 4\_PDFs\_with\_Errors
    - Duplicatas → 5\_PDFs\_with\_duplicates
    - PDFs que requerem OCR → 2\_PDFs\_to\_OCR
    - PDFs prontos para busca → 3\_PDFs\_to\_search
  4. Se houver arquivos que exigem OCR, o sistema perguntará se deseja realizar o Reconhecimento Óptico de Caracteres. Clique em Sim e aguarde a conclusão.
-



## Etapa 6: Realizando Buscas nos PDFs

1. Após o pré-processamento, volte ao menu principal.
  2. Clique em Busca para abrir o módulo de busca.
  3. Escolha entre os três tipos de busca:
    - Busca por Espécie — para buscar uma espécie específica.
    - Busca por Lista de Espécies — para buscar múltiplas espécies com base em uma planilha Excel (Data.xlsx).
    - Busca por Termos ou Frases — para buscar palavras-chave ou expressões.
  4. Configure os parâmetros conforme o tipo de busca selecionado e clique em Iniciar.
  5. Aguarde a geração dos relatórios.
- 

## Etapa 7: Visualizando e Exportando os Resultados

- Os relatórios são salvos automaticamente na pasta 6\_PDFs\_with\_target\_data.
  - Eles estarão organizados em subpastas conforme a espécie ou termo pesquisado.
  - Abra os arquivos .xlsx com o Microsoft Excel ou software compatível para análise detalhada.
- 

## Dicas Finais

- Sempre realize o pré-processamento antes de efetuar as buscas.
  - Mantenha suas pastas organizadas para evitar erros.
  - Consulte o FAQ para resolver problemas comuns.
  - Para suporte técnico, entre em contato: [sac@pteryx.com.br](mailto:sac@pteryx.com.br)
-