

READ-COOP SCE

European Cooperative with limited liability



Como usar Transkribus – 10 passos (ou menos)

Esse documento é uma tradução e adaptação do documento How To Use Transkribus in 10 Steps, disponibilizado no site <https://readcoop.eu/transkribus/howto/use-transkribus-in-10-steps/>. Ele foi desenvolvido pela Profa. Dra. Livia Borges Souza Magalhães e Profa. Dra. Lucia Furquim Werneck Xavier para fins exclusivamente didáticos.

Esse documento é um guia básico para uso do Transkribus. Ele fornece informações para o manuseio do aplicativo. Se precisar de instruções mais detalhadas sobre as funções do Transkribus, dê uma olhada em nossos “Guias de Como fazer”, disponíveis na Wiki do Transkribus: <https://transkribus.eu/wiki/>

Informações para baixar a versão desktop do Transkribus, ou para instalar a versão mais recente:

- <https://transkribus.eu/>

Confira a Wiki do Transkribus para mais informações ou para acessar o "Guias Como fazer":

- <https://transkribus.eu/wiki/>

Transkribus e a tecnologia por trás do aplicativo foram possíveis graças aos seguintes projetos e sites:

- <https://read.transkribus.eu/>
- <https://transcriptorium.eu/>
- <https://github.com/transkribus/>

Contato

- Equipe do Transkribus: email@transkribus.eu

INTRODUÇÃO

O Transkribus pode ser usado para vários propósitos, mas os mais importantes são:

- i. Transcrever documentos para edições com fins acadêmicos;
- ii. Criar dados de treinamento para alimentar o sistema de Reconhecimento de Texto Manuscrito ou, em inglês, *Handwritten Text Recognition*, (HTR) para que ele possa aprender a decifrar seus documentos históricos.
- iii. Executar o HTR em seus documentos e produzir transcrições geradas automaticamente.
- iv. Pesquisar por palavras distintas em suas coleções de documentos por meio da ferramenta **“Keyword Spotting”**, muito mais consistente do que as ferramentas de busca comuns aos textos.
- v. Articular a comunidade de interessados em estudos de textos manuscritos, uma vez que o aplicativo se aprimora através da ampliação do uso: quanto mais dados forem enviados ao Transkribus, mais eficiente será o aplicativo e, especialmente, o Reconhecimento de Texto Manuscrito.

O Transkribus foi produzido como uma infraestrutura de pesquisa pelo Projeto H2020 Project READ (Reconhecimento e Enriquecimento de Documentos de Arquivos <https://read.transkribus.eu/>).

Reserve algum tempo para explorar o Transkribus e familiarizar-se com o seu funcionamento. Para facilitar, existem "Guias de Como Fazer", em inglês, que apresentam instruções sobre as diferentes funções da Plataforma. Você pode encontrá-los no link: <https://transkribus.eu/wiki/>

COMO USAR TRANSKRIBUS – REGISTRE-SE NO SITE

Vá em: <http://transkribus.eu/> e leia o **“acordo de usuário”**:

https://transkribus.eu/Transkribus/docs/TranskribusTermsOfUse_v04-2016.pdf

É importante destacar que todos os documentos que você inserir no Transkribus são privados, ou seja, somente você, e mais ninguém, tem acesso a eles, a não ser que você compartilhe o acesso para terceiros.

A equipe do Transkribus segue rigorosamente todas as diretivas da União Europeia sobre a proteção de dados e privacidade. Respeitamos sua privacidade e

apenas usaremos os dados para melhorar nossos serviços e apoiar pesquisas em ciências humanas e informática!

PARA BAIXAR O APLICATIVO DO TRANSKRIBUS DA PÁGINA

Vá a página do Transkribus <http://transkribus.eu/> e clique em “Download”.

O Transkribus funciona em *Windows*, *MacOS* e *Linux*. Caso precise de ajuda para instalar a Plataforma, consulte a Wiki do Transkribus: https://transkribus.eu/wiki/index.php/Download_and_Installation

Para os usuários de *MacOS*, é possível que uma mensagem de erro apareça quando você abrir o Transkribus pela primeira vez. Para resolver esse problema aperte o botão direito do mouse para abrir a “Configuração do Sistema” e adicionar a exceção de segurança para o Transkribus.

Após baixar o Transkribus, não deixe de descompactar o arquivo. O aplicativo não pode ser iniciado a partir de um arquivo compactado (zipped file).

Abra o Transkribus e, depois, abra o aplicativo e use o botão “Login” na aba “Server” (servidor), conforme indicado na Figura 1.

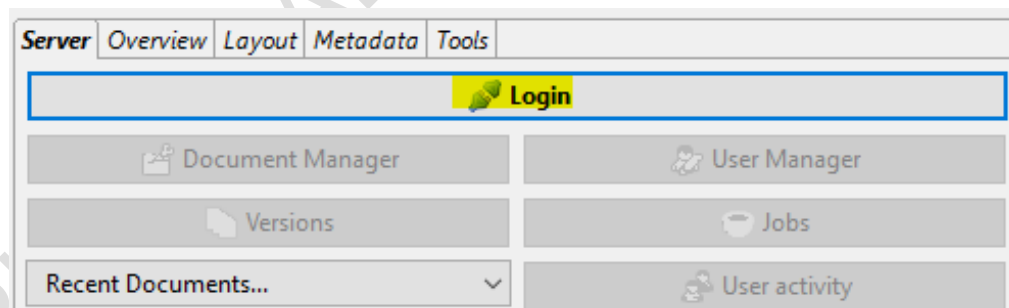


Figura 1 “Login”

Você terá acesso a sua coleção particular, nomeada com o seu email de acesso. Essa coleção é composta, inicialmente, por alguns documentos para você fazer testes de uso do Transkribus. Para acessá-los, clique no botão “Collections” (Coleções), na aba “Server”, conforme Figura 2.

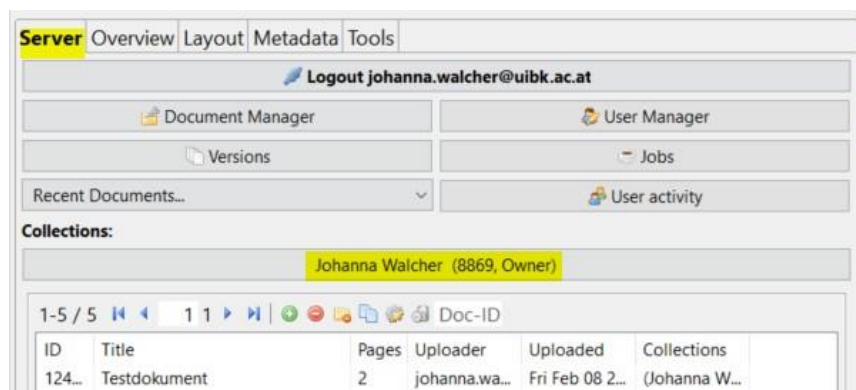


Figura 2 Documentos para teste na sua coleção

INSIRA SEUS DOCUMENTOS NO TRANSKRIBUS

O Transkribus permite que você trabalhe com seus próprios documentos, tanto no seu computador quanto transferindo-os para o servidor da Plataforma, ou seja, *online*. Entretanto, para usar os recursos automatizados, como o HTR, você precisará estar conectado na plataforma que, pela sua configuração, é capaz de processar arquivos em formato PDF, JPEG, PNG e TIFF. Arquivos JP2 não são suportados, infelizmente.

Você pode inserir documentos que você mesmo digitalizou ou usar o aplicativo **DocScan**, criado para *smartphones Android*. Com ele você pode tirar fotos e enviá-las diretamente para o Transkribus. Para mais informações: <https://scantent.cvl.tuwien.ac.at/en>

Você também pode baixar documentos da Internet e enviá-los para o Transkribus. Muitas bibliotecas e arquivos seguem as políticas de Acesso Aberto e, portanto, estão incentivando o uso livre de suas coleções, mas recomendamos que consulte as instituições diretamente e pergunte se é permitido colocar as imagens de seus documentos no Transkribus!

Para inserir documentos, clique no botão “Import document(s)”, indicado na Figura 3, para transferir as imagens de seu computador para a Plataforma. Nota: as imagens devem estar arquivadas numa pasta separada no seu computador antes de serem transferidas para o Transkribus.

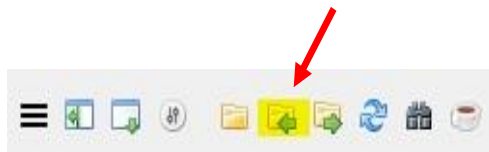


Figura 3 Botão "Import documentos" usado para inserir seus documentos no Transkribus

Você pode adicionar seus documentos a uma das suas coleções já existentes, apertando o botão "Add to collection" na janela "Document ingest/upload", conforme indicado na Figura 4.

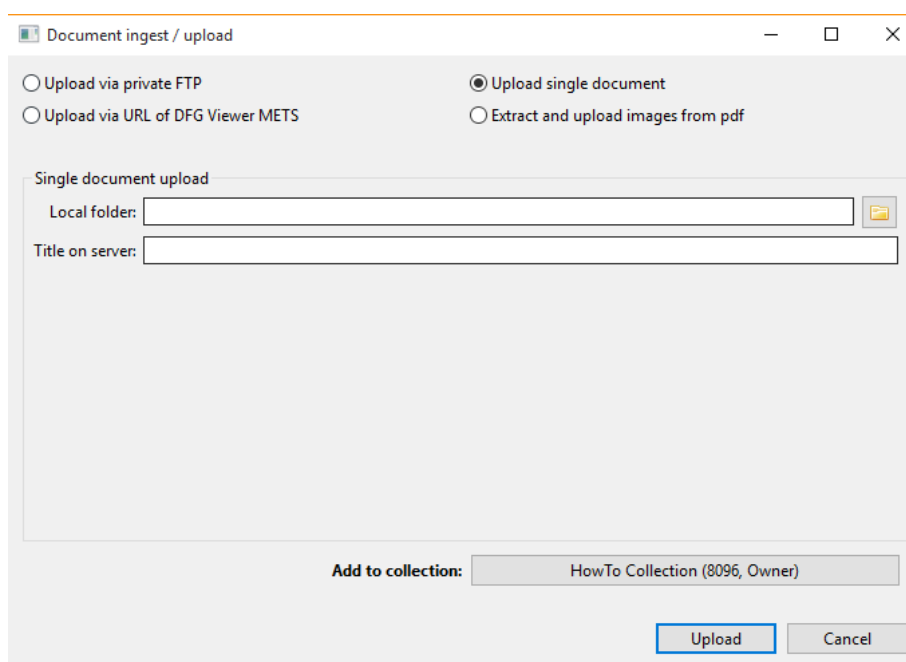


Figura 4 Acrescentar documento a uma coleção já existente

Ou você pode criar uma coleção nova, clicando no botão "Create" (FIGURA 5)

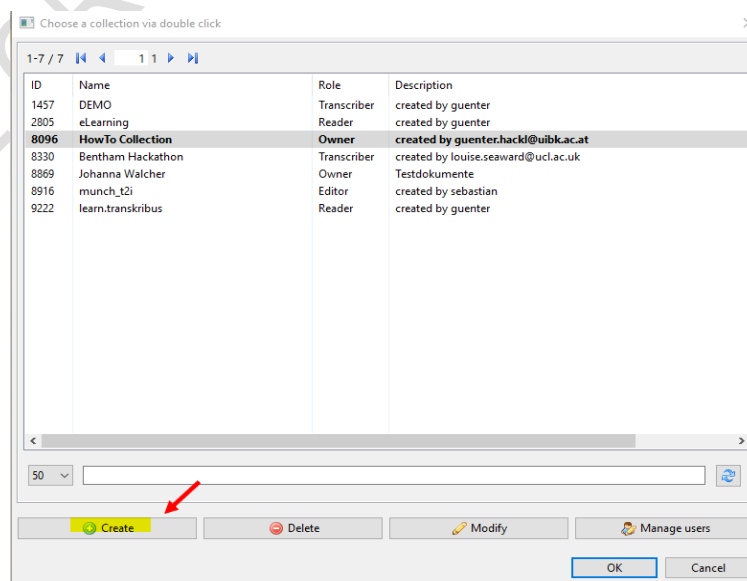


Figura 5 Botão "Create", usado para criar uma coleção nova

Para acessar seus documentos, clique no botão “Collections”, na aba “Server” e escolha a coleção desejada. Em seguida, ainda na aba “Server”, um pouco mais abaixo, clique duas vezes no documento que deseja abrir.

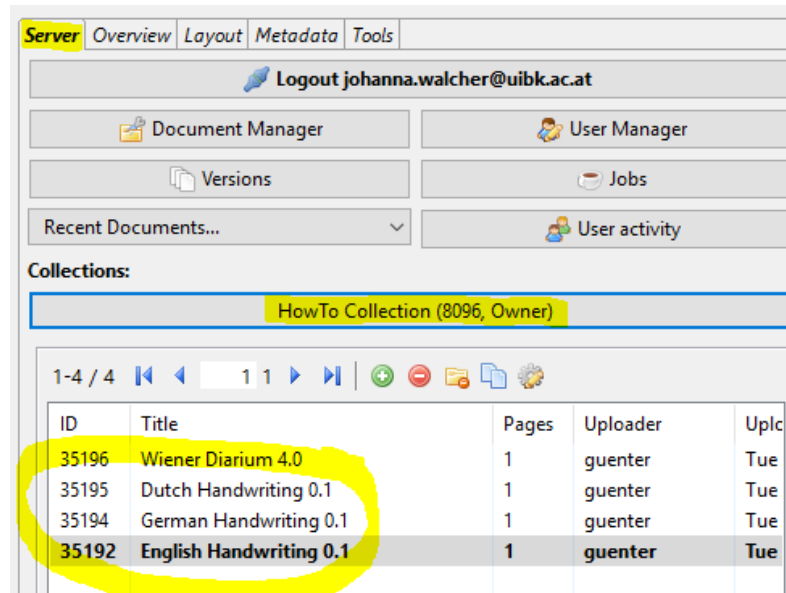


Figura 6 Captura de tela para acessar as coleções presentes no seu servidor

Como dito anteriormente, todos os documentos carregados no Transkribus são de uso privados, mas, você pode autorizar outras pessoas a verem seus documentos, caso deseje compartilhá-los. Na aba “Server”, use o botão “User Manager” para adicionar usuários à sua coleção. Você só pode dividir coleções com qualquer pessoa que possui uma conta no Transkribus, basta inserir o e-mail dela. Ressaltamos, inclusive, que se você estiver trabalhando em equipe, talvez seja mais fácil transcrever na *Interface da Web do Transkribus* disponível no link: <https://transkribus.eu/read>.

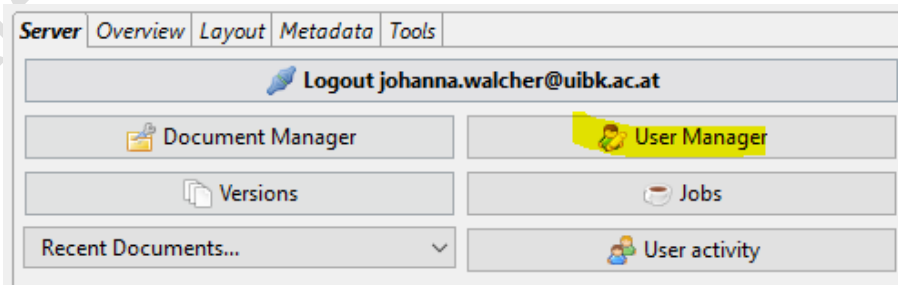


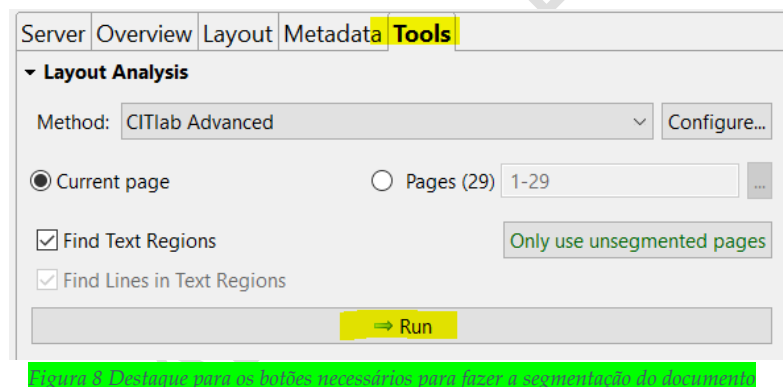
Figura 7 Botão “User Manager” para organizar o acesso a sua coleção

SEGMENTAR SEUS DOCUMENTOS EM LINHAS

Para alimentar a ferramenta de HTR com dados para treinamento, os documentos precisam ser segmentados em linhas. Isso pode ser feito automaticamente no Transkribus, basta ir na aba "Tools" (Ferramentas) e verificar se a opção "Find Text Regions" está selecionada. Se estiver, aperte "Run", conforme indicado na Figura 8.

Você pode escolher entre segmentar a página atual "current page", diversas páginas selecionadas ou o documento inteiro.

Após apertar o "run" as regiões de texto e as linhas do seu documento serão identificadas automaticamente, estando pronto para iniciar o processo de treinamento do HTR, por meio da transcrição.



COMECE SUA TRANSCRIÇÃO

Quando você abrir o documento segmentado, verá marcações de linhas. Elas são chamadas de linhas de base (*baselines* em inglês). É por meio dessa marcação que você deve se guiar para fazer a transcrição, que deve ser feita no campo "Text Editor" (Editor de Texto).

Clique no botão "Viewing Profiles" (Modo de visualização), conforme indicado na imagem a seguir, e selecione a opção "Transcription". Para cada "baseline", haverá uma linha correspondente no campo **Text Editor**. Transcreva o texto linha por linha, exatamente como aparece na imagem.

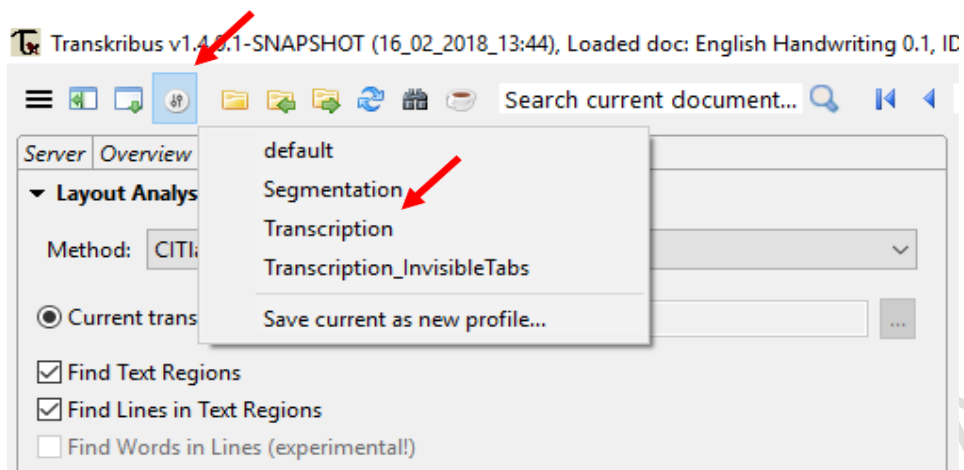


Figura 9 Modo Transcrição

Caso o seu documento apresente caracteres especiais, você pode inseri-los usando o “**Virtual Keyboards**” (Teclado Virtual) na barra de ferramentas do Text Editor, conforme Figuras 10 e 11.



Figura 10 Botão “Virtual Keyboards”

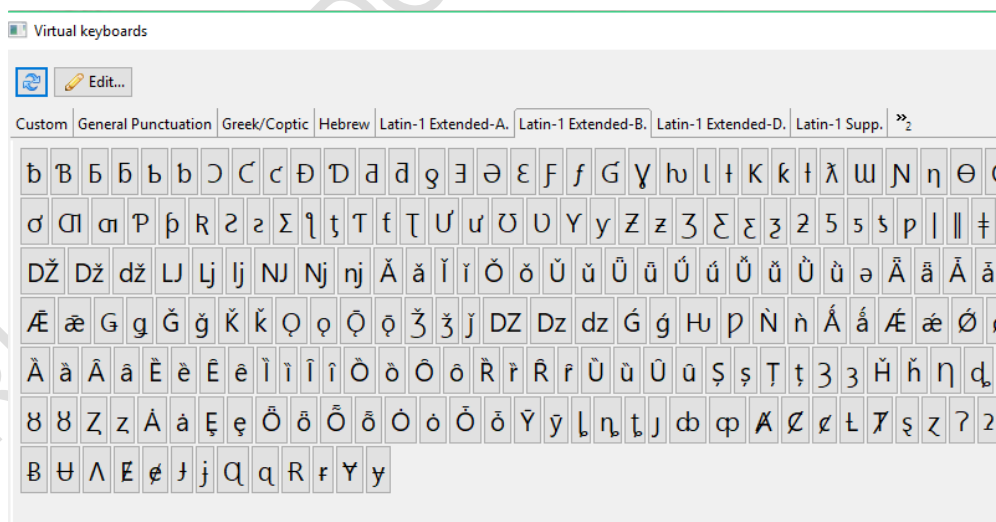


Figura 11 Teclado Virtual com caracteres especiais

SALVAR E EXPORTAR A SUA TRANSCRIÇÃO

Feita a transcrição, não esqueça de salvá-la. Para isso, basta apertar o botão “Save” no menu principal, destacado na figura 12:

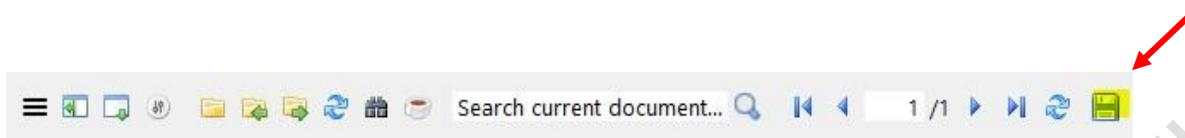
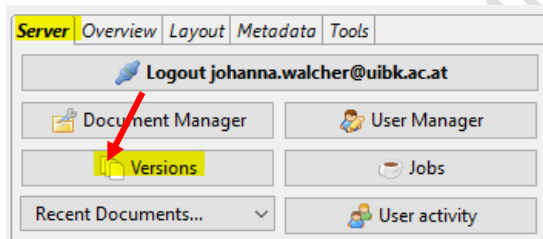


Figura 12: Destaque no botão “Save”

Após salvar, se você clicar no botão “Versions” na guia “Server” (FIGURA 13), poderá verificar que uma nova versão foi criada. Isso significa que, sempre que necessário, você poderá acessar versões anteriores do documento

Figura 13 - Clique no botão “Version” para acessar as versões anteriores do seu documento



É importante informar que você pode exportar todo o documento transcrito (ou parte dele) usando o botão “Export document”.



Figura 14 Botão “Export document”

O botão abrirá uma janela em que você poderá escolher se fará a exportação do documento inteiro, da sua coleção, o formato de saída do documento, que pode ser PDF, TEI, DOCX, TXT, e outras informações que podem ser selecionadas dependendo do seu objetivo com a transcrição como, por exemplo, se a transcrição deverá ser acompanhada das imagens. A versão exportada será enviada para o e-mail que você cadastrou na sua conta do Transkribus.

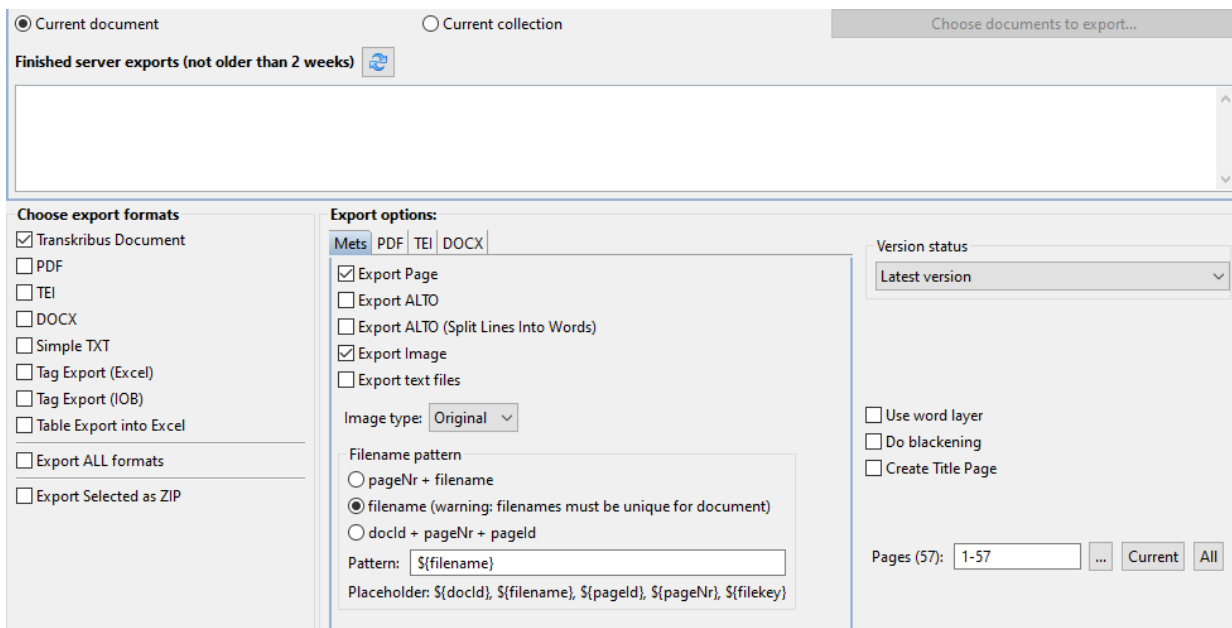


Figura 15 Janela de exportação de documento

USE RECONHECIMENTO DE TEXTO MANUSCRITO (HTR) EM SEUS DOCUMENTOS

É simples ter seus documentos reconhecidos pelo computador. Você pode começar a treinar um modelo com cerca de 5.000 palavras transcritas de texto impresso ou 15.000 palavras de texto manuscrito. Para iniciar o processo de treinamento, envie para a plataforma Transkribus um e-mail depois de segmentar e transcrever o primeiro lote de páginas (email@transkribus.eu). Nesse e-mail você deve informar **XXXXX**

Você receberá da plataforma a permissão para treinar seu próprio modelo. Se precisar de mais informações sobre isso, consulte o guia "Como treinar um modelo", disponível no link: **XXXXXXXXXXXXXXXXXXXX**

Uma vez que um modelo HTR foi treinado para seus documentos, ele pode ser aplicado através do botão "Run" na seção "Text Recognition" na guia "Tools". Você pode selecionar uma ou mais páginas de seus documentos e iniciar o reconhecimento, conforme indicado na Figura 16.

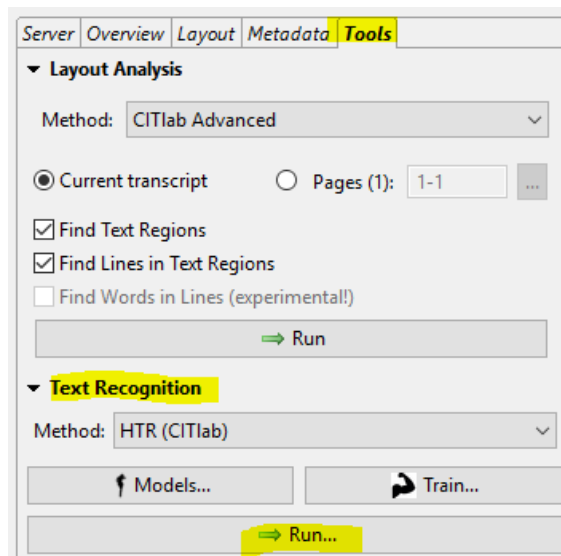


Figura 16 Executar o HTR

Se você clicar em “Run” e depois em “Configure”, verá informações sobre o seu modelo.

No lado esquerdo da janela você pode ter uma visão geral dos modelos disponíveis. Já na parte superior direita da janela, serão expostos os detalhes do modelo. Na parte inferior direita aparece um gráfico indica a precisão do seu modelo com a Taxa de Erro de Caractere (CER), ou seja, a porcentagem de caracteres que foram transcritos incorretamente pelo HTR. A linha azul representa o andamento do treinamento. A linha vermelha representa o andamento das avaliações no conjunto de teste de dados que foi reservado durante o processo de treinamento (FIGURA 17).

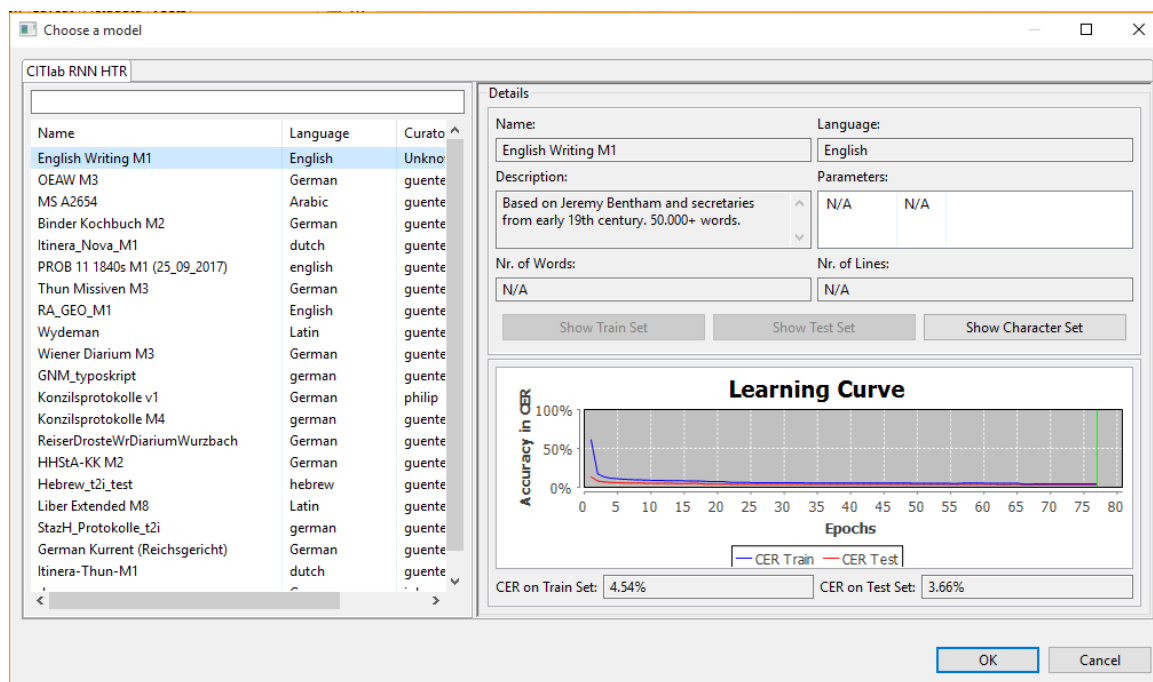


Figura 17 Visão geral do modelo e curva de aprendizado

Depois que o HTR terminar, os resultados aparecerão diretamente em uma nova versão do seu documento no Transkribus. É possível avaliar a precisão da transcrição automática usando o “Compute Accuracy”, uma função na aba “Tools”, como indicado na Figura 18.

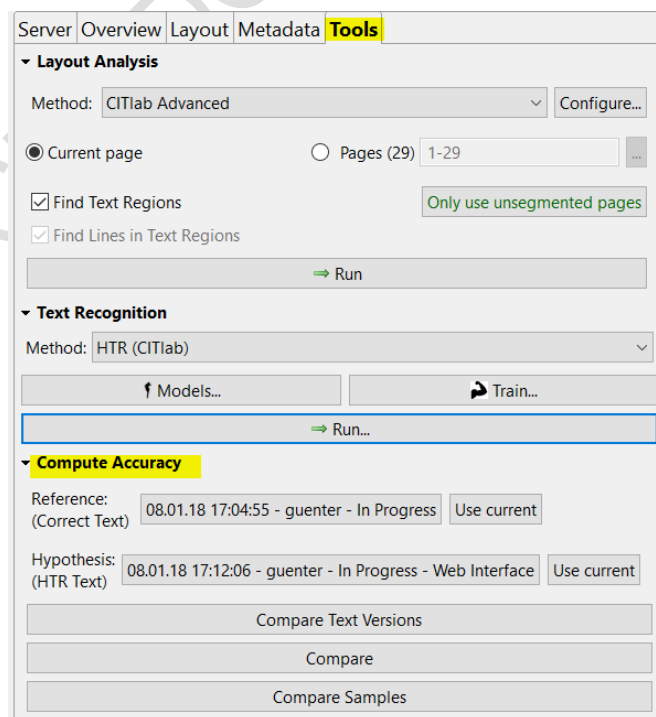


Figure 18 Compute the accuracy of the HTR

LOCALIZAÇÃO DE PALAVRAS-CHAVE

Assim que tiver um modelo HTR para seus documentos, você poderá pesquisá-los com a função de localização de palavras-chave.

Primeiro, execute um modelo HTR em seus documentos para produzir a transcrição automática e, depois, clique no botão do binóculo mostrado na Figura 19.



Figure 19 Open the "Search for..." window to use the Keyword Spotting function

O aplicativo abrirá uma janela. Nela você deve escolher a guia "KWS" e digitar a palavra que deseja pesquisar na caixa "Palavra-chave 1" (FIGURA 20). Feito isso, pressione o botão "Search" e uma janela de confirmação aparecerá (FIGURA 21). Clique em "Yes" para iniciar sua consulta de localização das palavras-chave.

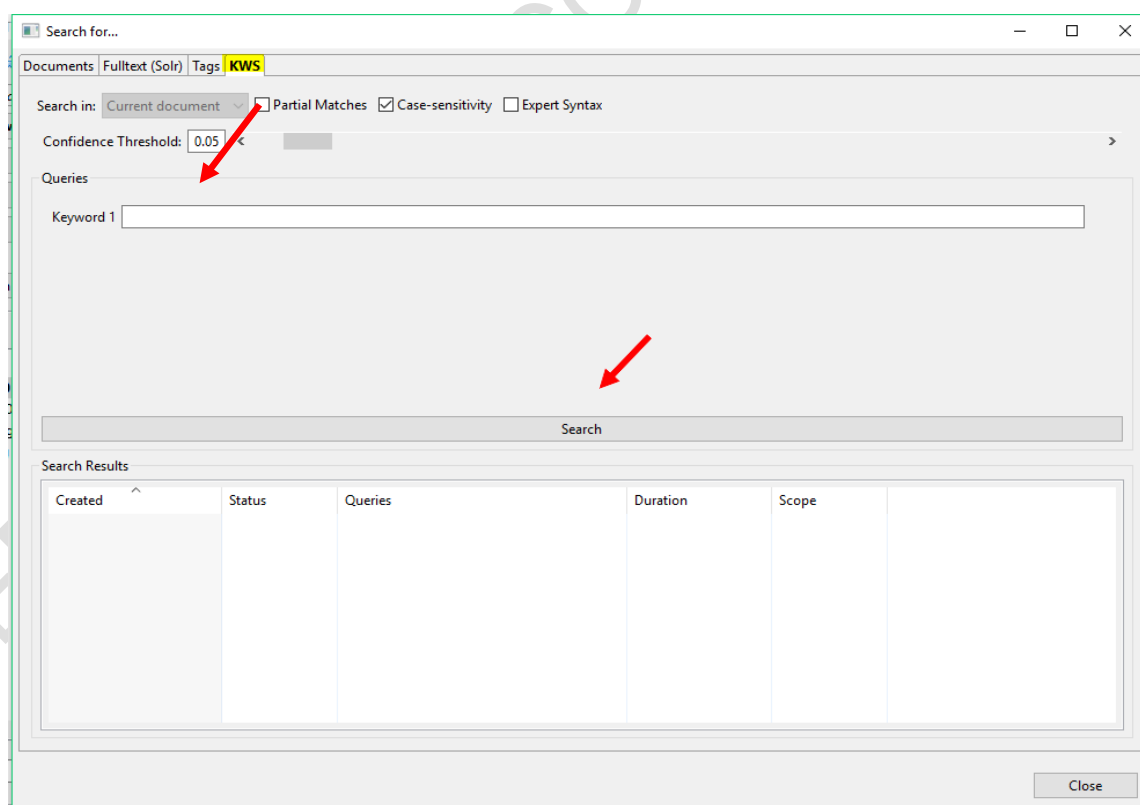


Figure 20 Janela para usar a função de localização de palavras-chave

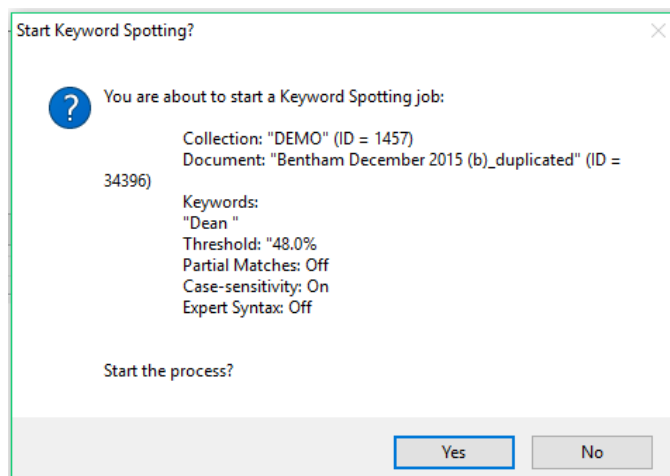


Figura 21 Janela de confirmação

Assim que sua consulta for concluída, clique duas vezes na data e no valor numérico no Coluna "Created" para acessar seus resultados de pesquisa, conforme indicado na Figura 22.

Created	Status	Queries	Duration	Scope
02.01.18 12:52:07	Completed	"Dean " (10)	7.45 sec.	Document 343...
02.01.18 10:53:22	Completed	"William" (3)	5.07 sec.	Document 343...
28.12.17 13:46:46	Completed	"dean" (117), "and" (333)	15.88 sec.	Document 343...
28.12.17 13:46:42	Completed	"dean" (117), "and" (333)	10.45 sec.	Document 343...
28.12.17 13:44:10	Completed	"dean" (117)	7.31 sec.	Document 343...

Figura 22 Resultado da busca de palavras-chave

A janela "Keyword Spotting Results" mostrará uma lista de locais onde essa palavra-chave aparece.

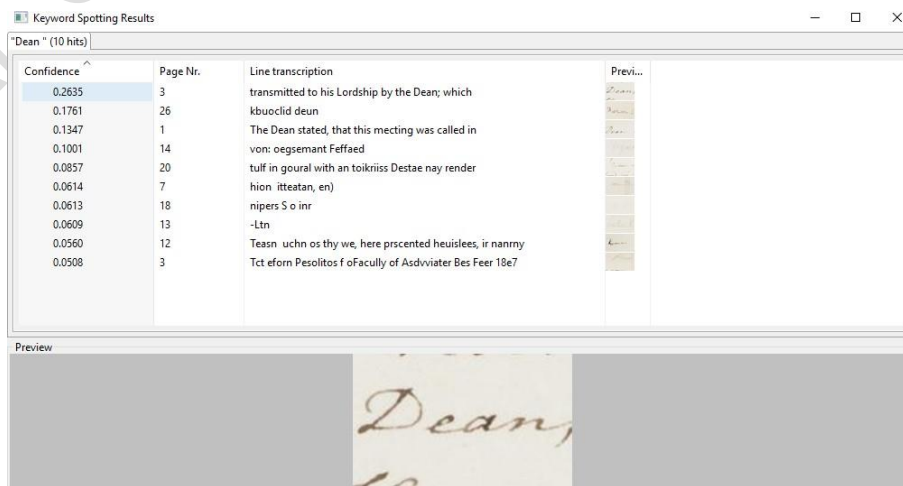


Figura 23 Informações sobre os resultados do Spotting de palavras-chave