



**Recognition and Enrichment
of Archival Documents**



How To use Transkribus – in 10 steps (or less)

*Version v1.4.0. (22_02_2018_15:07)
Last update of this guide 26.09.2019*

This document is a basic introduction to Transkribus. It provides a simple standard workflow for working with the platform. If you need more detailed instructions on the functions of Transkribus please have a look at our How to Guides, which can be found on the Transkribus Wiki: <https://transkribus.eu/wiki/>

Download the Transkribus Expert Client, or make sure you are using the latest version:

- <https://transkribus.eu/>

Consult the Transkribus Wiki for further information and other How to Guides:

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

Transkribus and the technology behind it are made available via the following projects and sites:

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

Contact

- The Transkribus Team: email@transkribus.eu



The READ project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 674943.

1. Introduction

- a. Transkribus can be used for several purposes. The most important are:
 - i. Transcribe documents for a scholarly edition
 - ii. Create training data to feed the **Handwritten Text Recognition (HTR)** system so it can learn to decipher your historical documents.
 - iii. Run HTR on your documents and receive **automatically generated transcripts**.
 - iv. Search for distinct words in your document collections **with Keyword Spotting** which is much more powerful than standard full-text search.
 - v. The platform lives from the community. The more data uploaded to Transkribus, the more efficient the program and especially the Handwritten Text Recognition will get.
- b. Transkribus is offered as a research infrastructure by the H2020 Project **READ (Recognition and Enrichment of Archival Documents)** <https://read.transkribus.eu/>).
- c. Take some time to explore Transkribus and become familiar with how it works. To make it easier we have created several How to Guides, which give instructions on the different functions of the platform. You can find them on the Transkribus Wiki: <https://transkribus.eu/wiki/>

2. To use Transkribus - register at the website

- a. Go to: <http://transkribus.eu/>
- b. Read our **user agreement**: https://transkribus.eu/Transkribus/docs/TranskribusTermsOfUse_v04-2016.pdf
- c. All documents uploaded to Transkribus are “private”, which means that no one except you has access to them.
- d. The Transkribus team fully supports all EU directives on data protection and privacy. We will respect your privacy and only use the data to improve our services and support research in humanities and computer science!

3. Download Transkribus from the website

- a. Go to the Transkribus website <http://transkribus.eu/> and click “Download”.
- b. Transkribus runs on Windows, MacOS and Linux. If you need help installing the platform, consult the Transkribus wiki: https://transkribus.eu/wiki/index.php/Download_and_Installation
- c. If you use MacOS an error message may appear when you try to open Transkribus for the first time. To remedy this:
 - i. right click the Track Pad to open the Context Menu and add a security exception for Transkribus.
- d. Once you have downloaded Transkribus, make sure you unzip the file. The program cannot be started from the zipped file!

4. Open Transkribus

- a. Start the tool and use the “Login” button in the “Server” tab.

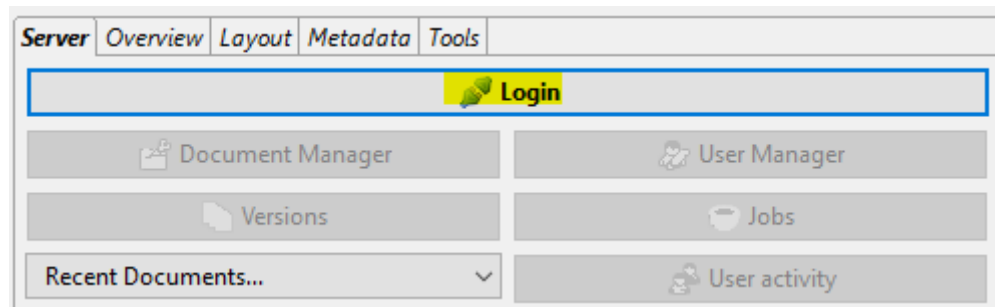


Figure 1 Login

- b. You will have access to your private collection named after your email address). This collection includes some test documents that you can experiment with.
- c. You can find it by clicking the “Collections” button in the “Server” tab.

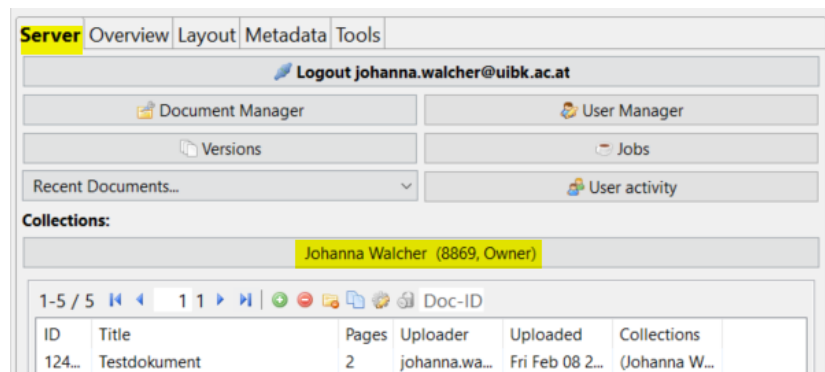


Figure 2 Test documents in your collection

5. Upload your documents

- a. Transkribus allows you to work with your own documents, either locally or by uploading them to the server.
- b. Automated processes can only be performed if the documents are uploaded to the Transkribus platform. The platform can process PDF, JPEG, PNG and TIFF files. JP2 files are not supported unfortunately.
- c. You can upload documents which you have scanned yourself. You can also use our **DocScan app** for Android smartphones to take images and upload them directly to Transkribus. For more information: <https://scantent.cvl.tuwien.ac.at/en>
- d. You may also download documents from the Internet and upload them to Transkribus. Many libraries and archives follow Open Access policies and are therefore encouraging further usage of their collection – you can ask archives and libraries directly if you can upload images of their documents to Transkribus!
- e. Click the “Import document(s)” button to transfer the images from your computer to the platform. Note: the images need to reside in a separate folder on your computer before you upload them to Transkribus!

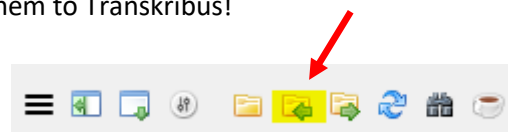


Figure 3 Upload your documents to Transkribus

- f. You can add your documents to one of your existing collections or create a new one by clicking the “Add to collection” button at the bottom of the “Document ingest/upload” box and then clicking “Create”.

Document ingest / upload

☐ Upload via private FTP
 ☒ Upload single document

☐ Upload via URL of DFG Viewer METS
 ☐ Extract and upload images from pdf

Single document upload

Local folder:

Title on server:

Add to collection: HowTo Collection (8096, Owner)

Upload Cancel

Figure 4 Add documents to one of the existing collections or create a new one

Choose a collection via double click

1-7 / 7

ID	Name	Role	Description
1457	DEMO	Transcriber	created by guenter
2805	eLearning	Reader	created by guenter
8096	HowTo Collection	Owner	created by guenter.hackl@uibk.ac.at
8330	Bentham Hackathon	Transcriber	created by louise.seaward@ucl.ac.uk
8869	Johanna Walcher	Owner	Testdokumente
8916	munch_t2i	Editor	created by sebastian
9222	learn.transkribus	Reader	created by guenter

50

Create Delete Modify Manage users

OK Cancel

Figure 5 Create your own collection

- g. To access your documents, click on the “Collections” button in the “Server” tab and choose your collection. Then double-click on the documents in the box at the bottom of the “Server” tab to open them.

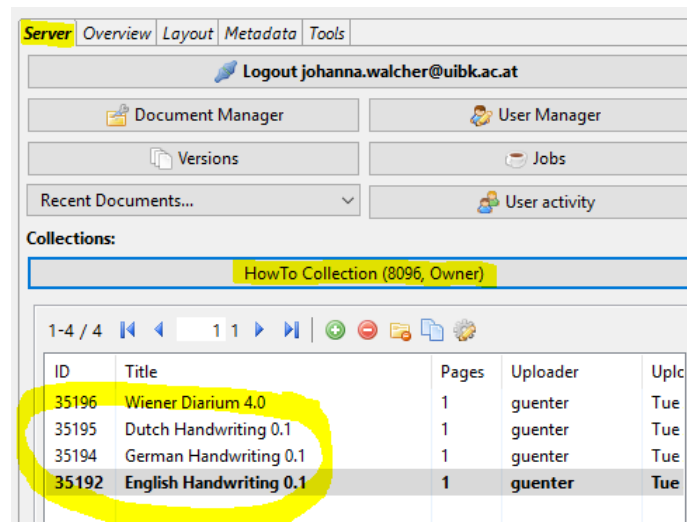


Figure 6 Open the documents in your collection

- h. **All documents uploaded to Transkribus are private by default.** You can give other users authorisation to view your documents if you wish. Use the “User Manager” button in the “Server” tab to add users to your collection. You can only share collections with users who have a Transkribus account.

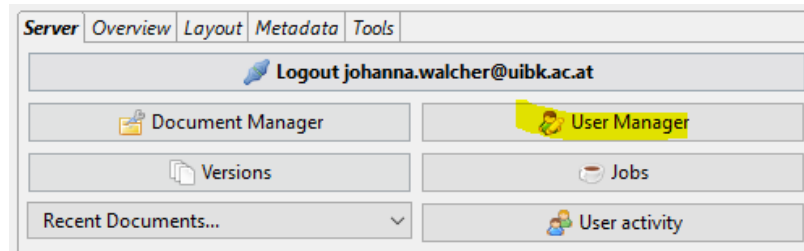


Figure 7 “User Manager” button for managing access to your collection

6. Segment your documents into lines

- In order to be able to feed the HTR engine with training data the documents need to be segmented into lines. This can be done automatically in Transkribus.
- Open the “Tools” tab.
- Make sure “Find Text Regions” is selected and press “Run”.
- You can choose to segment the current page or a batch of pages.
- The lines and text regions in your document will be detected automatically.

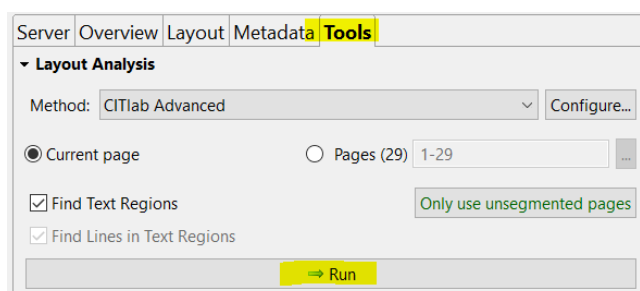


Figure 8 Segmentation

7. Start your transcription

- Once the baselines are visible on your image you can write text into the Text Editor field.
- Click on the “Viewing Profiles” button and select the “Transcription” view.
- For each baseline, there will be **a corresponding line in the Text Editor**. Transcribe the text line by line, exactly as it appears in the image.

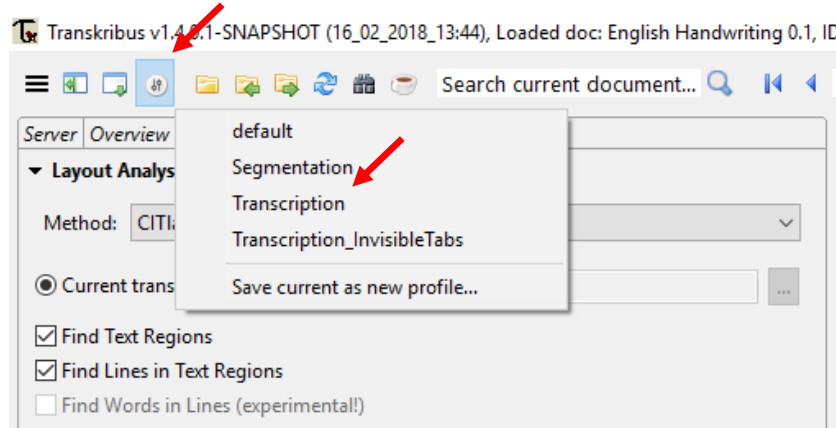


Figure 9 Transcription view

- Special characters can be found in the “**Virtual Keyboards**” button in the Text Editor toolbar.

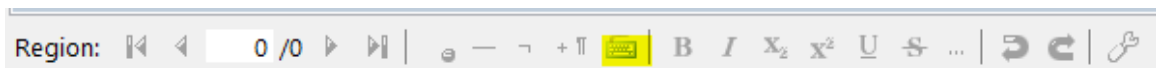


Figure 10 “Virtual Keyboards” button

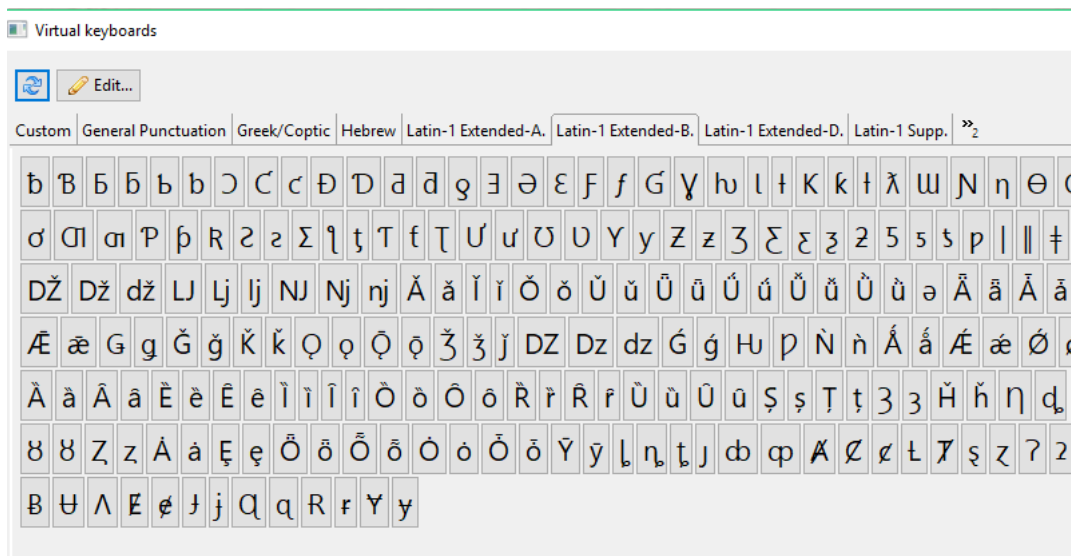


Figure 11 Virtual keyboards

- If you are working in a team, you might find it easier to transcribe in **the Transkribus Web Interface**. This is a lite version of Transkribus which is simple to use: <https://transkribus.eu/read>

8. Save and export your transcription

- Press the “Save” button in the Main Menu to save the document in Transkribus.

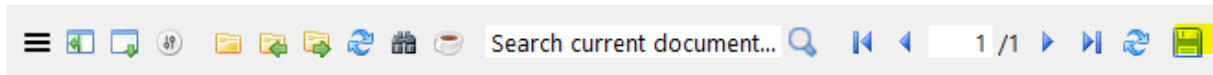


Figure 12 Saving the changes in your document

- If you click on the “Versions” button in the “Server” tab, you will see that a **new version** has been created. This means that you can always access previous versions of a document should you need to.

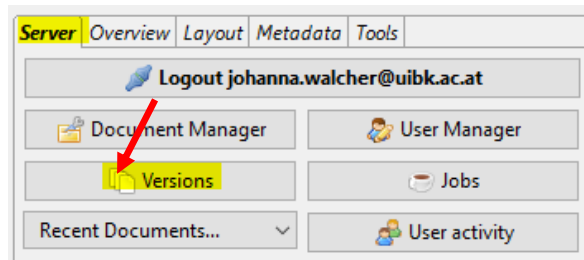


Figure 13 Click the “Versions” button to access previous versions of your document

- You can also **export the whole document** at any point of the process by clicking the “Export document” button.



Figure 14 “Export document” button

9. Use Handwritten Text Recognition (HTR) on your documents

- It is simple to have your documents recognised by the computer. You can start training a model with around 5,000 transcribed words of printed text or 15,000 words of handwritten text. To start the training process please drop us a short email once you have segmented and transcribed a first batch of pages (email@transkribus.eu).
- You will receive the permission to train your own model from us. If you need more information on that please check the [How to Train a Model](#) guide.
- Once an **HTR model** has been trained for your documents, it can be applied via the “Run” button in the “Text Recognition” section in the “Tools” tab. You can select one or more pages of your documents and start recognition.

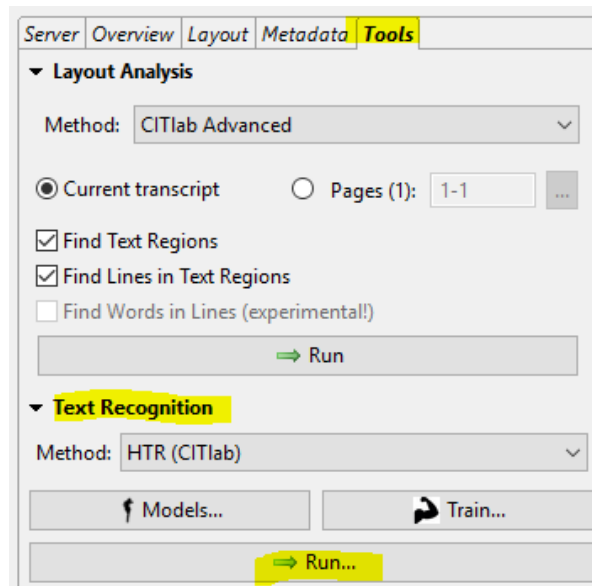


Figure 15 Run Handwritten Text Recognition

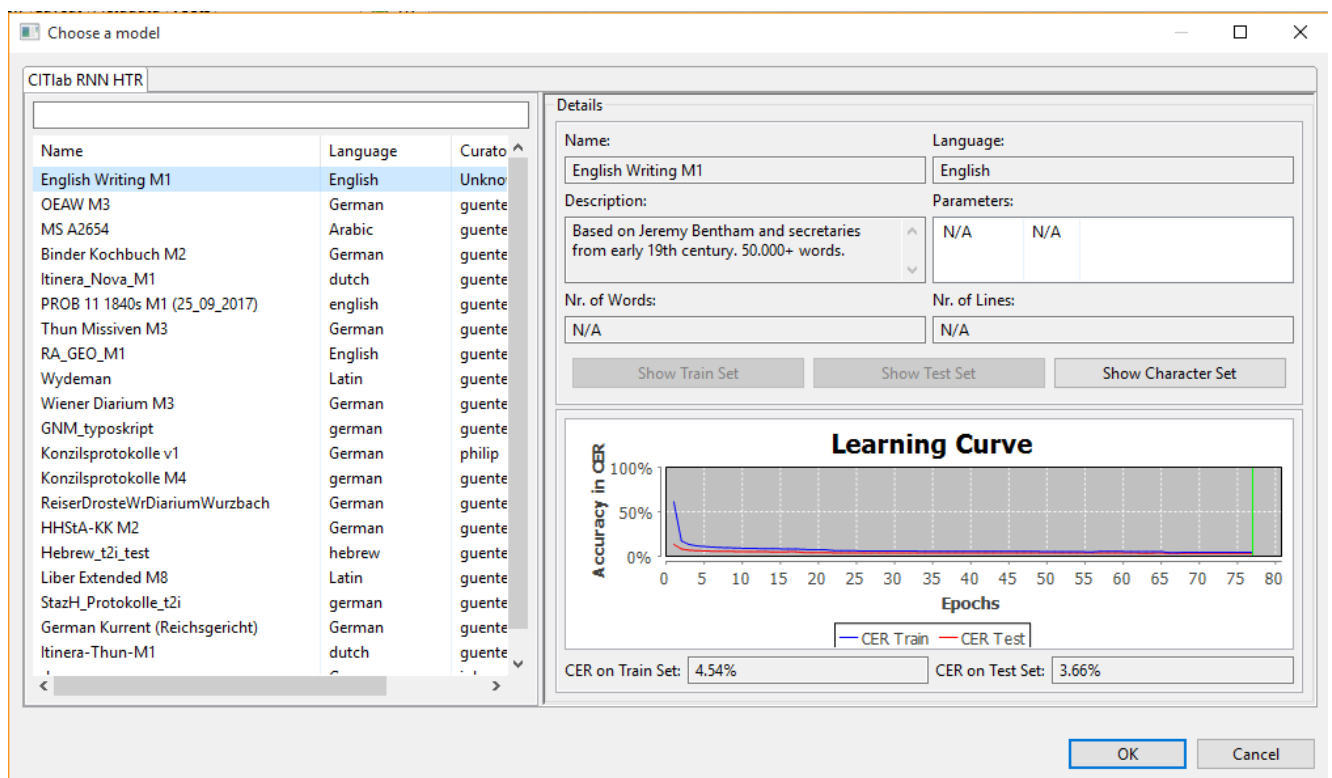


Figure 16 Model overview and learning curve

- d. If you click “Run” and then “Configure”, you will see information about your model.
- e. On the left side of the window you can see an overview of the available models.
- f. On the top right side of the window the details of the model are shown.
- g. The graph on the bottom right signifies the accuracy of your model with the Character Error Rate (CER), i.e. the percentage of characters that have been transcribed incorrectly by HTR. The **blue line** represents the progress of the training. The **red line** represents the progress of evaluations on the Test Set of data which was set aside during the training process.

- h. After the HTR has finished the results will appear directly on a new version of your document within Transkribus. It is possible to **evaluate the accuracy** of the automatic transcription using the “Compute Accuracy”-function in the “Tools” tab.

The screenshot shows the 'Tools' tab in the Transkribus interface. It is divided into three main sections: 'Layout Analysis', 'Text Recognition', and 'Compute Accuracy'. The 'Compute Accuracy' section is highlighted with a yellow background. It contains fields for 'Reference: (Correct Text)' and 'Hypothesis: (HTR Text)', each with a dropdown menu and a 'Use current' button. Below these fields are three buttons: 'Compare Text Versions', 'Compare', and 'Compare Samples'.

Figure 17 Compute the accuracy of the HTR

10. Keyword Spotting

- a. Once you have a HTR model for your documents, you will be able to search them with the Keyword Spotting function.
- a. First, run a HTR model on your documents to produce an automatic transcript.
- b. The Keyword Spotting function can be opened with the binoculars button shown in Figure 18.



Figure 18 Open the “Search for...” window to use the Keyword Spotting function

c. In the window which opens up choose the “KWS” tab.

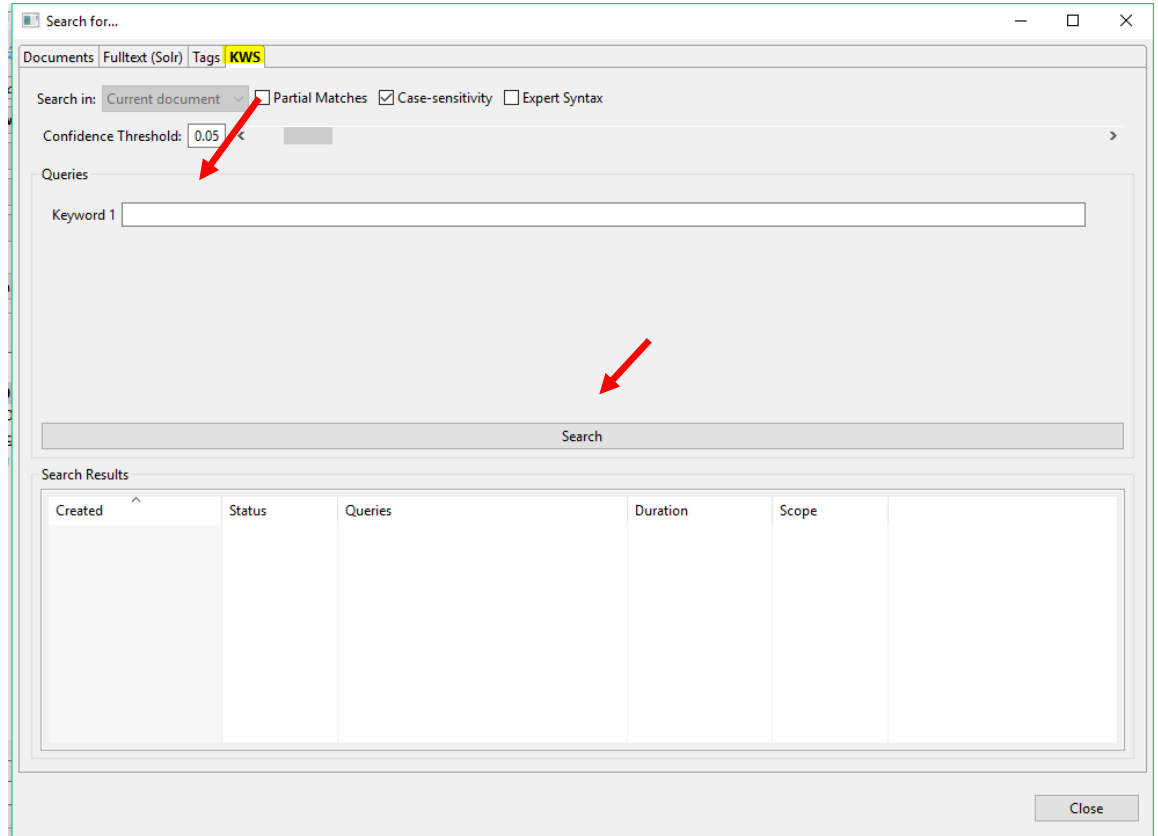


Figure 19 Window to use the Keyword Spotting function

- Simply type the word you would like to search for in the “Keyword 1” box and press the “Search” button.
- A confirmation window will pop-up. Click “Yes” to start your Keyword Spotting query

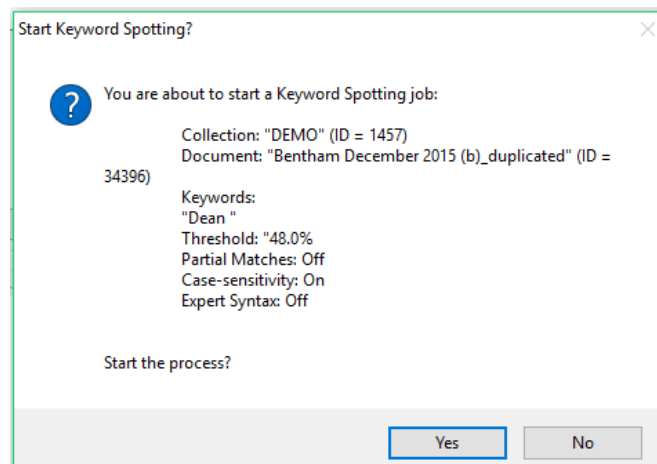


Figure 20 Confirmation window

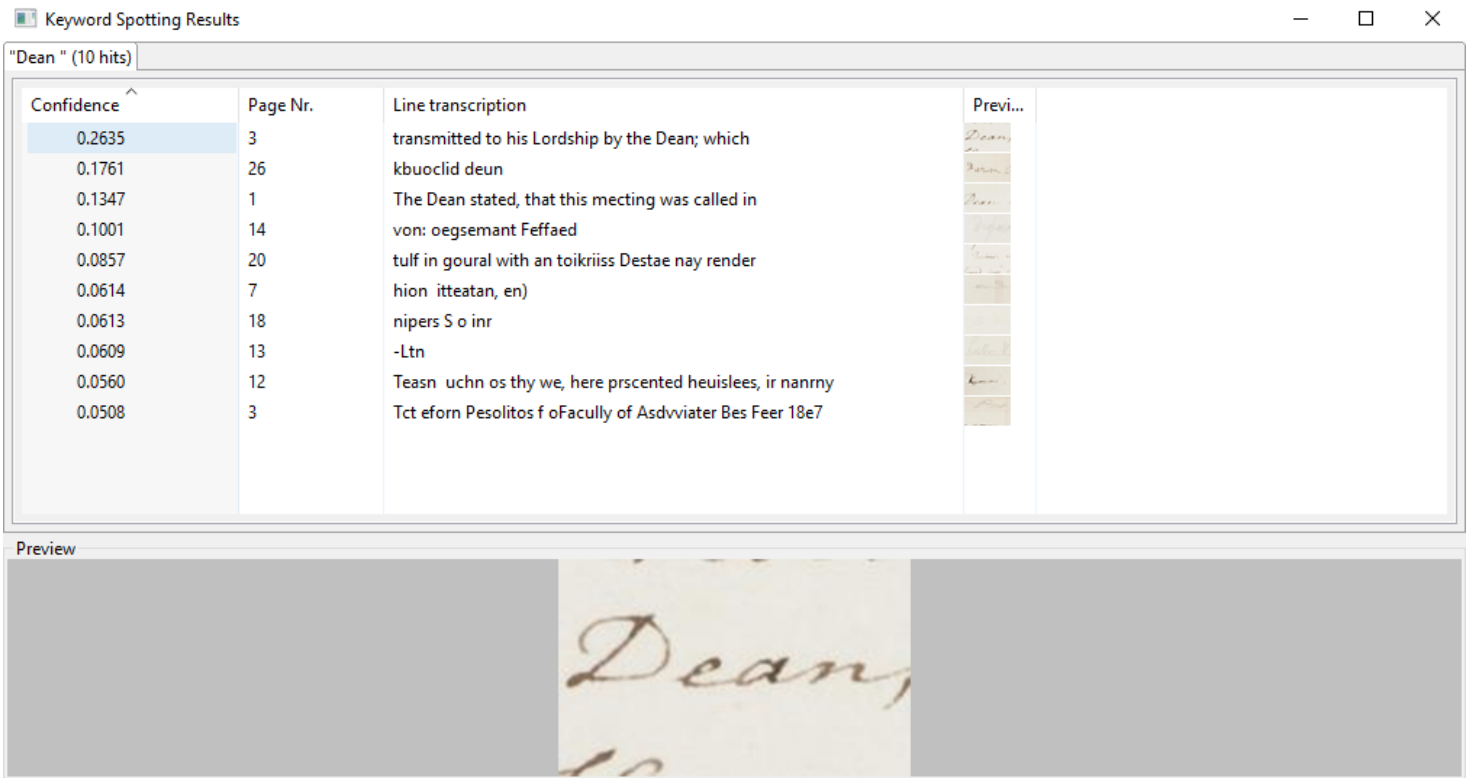
- Once your search query is finished double-click the date and numerical value in the “Created” column to access your search results



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...

Figure 21 Keyword Spotting results

- The “Keyword Spotting Results” window will show you a list of places where that keyword appears.



Confidence	Page Nr.	Line transcription	Previ...
0.2635	3	transmitted to his Lordship by the Dean; which	
0.1761	26	kbuoclid deun	
0.1347	1	The Dean stated, that this meeting was called in	
0.1001	14	von: oegsemant Feffaed	
0.0857	20	tulf in goural with an toikriiss Destae nay render	
0.0614	7	hion itteatan, en)	
0.0613	18	nipers S o inr	
0.0609	13	-Ltn	
0.0560	12	Teasn uchn os thy we, here prsented heuislees, ir nanrny	
0.0508	3	Tct eform Pesolitos f oFaculty of Asdvviater Bes Feer 18e7	

Preview

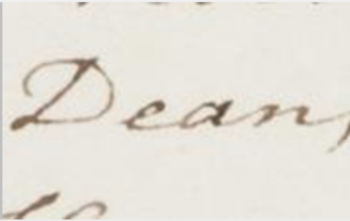


Figure 22 Information about your Keyword Spotting results

Credits

We would like to thank the many users who have contributed their feedback to help improve the Transkribus software.

Transkribus is made available to the public as part of H2020 e-Infrastructure Project READ (Recognition and Enrichment of Archival Documents) which received funding from the European Commission.