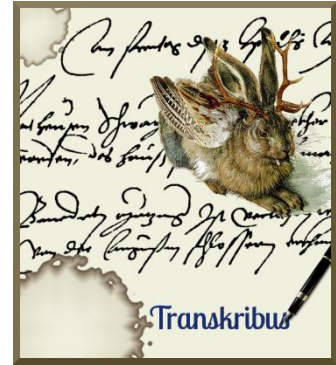


READ-COOP SCE

Europäische Genossenschaft mit beschränkter Haftung



Transkribus in 10 (oder weniger) Schritten

Version v1.4.0. (22_02_2018_15:07)

Letzte Überarbeitung dieser Anleitung 26.09.2019

Dieses Dokument ist eine grundlegende Einführung in die Transkribus Plattform. Sie finden hier einen einfachen Standard-Workflow. Wenn Sie detailliertere Informationen benötigen, ziehen Sie bitte die weiteren Transkribus Anleitungen zu Rate. Diese sind auf der Transkribus Wiki Seite zu finden: <https://transkribus.eu/wiki/>

Laden Sie den Transkribus Expert Client herunter oder stellen Sie sicher, dass Sie die neueste Version verwenden:

- <https://transkribus.eu/>

Besuchen Sie das Transkribus Wiki für mehr Informationen und weitere How to Guides:

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

Transkribus und die zugrundeliegende Technologie werden durch die folgenden Projekten und Plattformen verfügbar gemacht:

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

Kontakt

- Das Transkribus Team: email@transkribus.eu



Die Transkribus Plattform wird von der Europäischen **Genossenschaft READ-COOP SCE** zur Verfügung gestellt.

Bis Ende Juni 2019 wurde Transkribus im Rahmen des READ Projekts in der H2020 Förderlinie unter dem Vertrag 674943 finanziert.

1. Einführung

- a. Transkribus kann für verschiedene Zwecke genutzt werden. Die wichtigsten sind:
 - i. Dokumente für eine wissenschaftliche Edition transkribieren
 - ii. Trainingsdaten für die automatische Handschriftenerkennung (Handwritten Text Recognition – HTR) generieren. Damit lernt das Programm Ihre historischen Dokumente zu entziffern
 - iii. Die Handschriftenerkennung für Ihre Dokumente nutzen und automatisch generierte Transkripte bekommen
 - iv. Mit dem „Keyword Spotting“ nach bestimmten Wörtern in Ihren Kollektionen suchen - diese Funktion ist effektiver als Standard Volltext Suchen
 - v. Die Plattform lebt von der Gemeinschaft - umso mehr Daten in die Transkribus Plattform geladen werden, umso effizienter wird die Handschriftenerkennung
- b. **Transkribus** wird durch das H2020 Projekt **READ (Recognition and Enrichment of Archival Documents)** finanziert (<https://read.transkribus.eu/>).
- c. Nehmen Sie sich bitte etwas Zeit, um die Plattform zu entdecken und mit ihr vertraut zu werden. Um das noch einfacher zu machen, stellen wir verschiedene Anleitungen zur Verfügung, die Erklärungen zu den verschiedenen Transkribus Funktionen geben. Sie sind auf der Transkribus Wiki Seite zu finden: <https://transkribus.eu/wiki/>

2. Um Transkribus zu nutzen, registrieren Sie sich auf der Transkribus Website

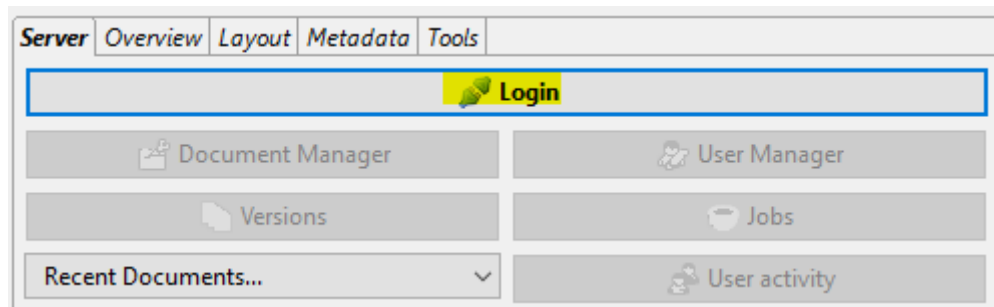
- a. <http://transkribus.eu/>
- b. Lesen sie unsere Nutzungsvereinbarungen: https://transkribus.eu/Transkribus/docs/TranskribusTermsOfUse_v04-2016.pdf
- c. Alle Dokumente, die in die Transkribus Plattform geladen werden, sind „privat“. Das bedeutet, dass niemand außer Sie auf die Dokumente Zugriff hat.
- d. Das Transkribus Team unterstützt alle EU Richtlinien zu Datenschutz und Privatsphäre. Wir respektieren Ihre Privatsphäre und benutzen die Daten deshalb nur, um unsere Services zu verbessern und die Forschung in den Geisteswissenschaften und der Informatik zu unterstützen.

3. Downloaden Sie Transkribus von unserer Website

- a. Gehen Sie auf die Transkribus Website (<http://transkribus.eu/>) und klicken Sie auf „Download“.
- b. Transkribus funktioniert mit Windows, MacOS und Linux. Falls Sie Hilfe bei der Installation benötigen, konsultieren Sie bitte die Transkribus Wiki Seite: <https://transkribus.eu/wikiDe/index.php/Hauptseite#Installation>
- c. Wenn Sie mit MacOS arbeiten kann es passieren, dass eine Fehlernachricht erscheint, sobald Sie Transkribus das erste Mal zu öffnen versuchen. Um das zu umgehen:
 - i. Machen Sie einen Rechtsklick auf das „Track Pad“ um das „Context Menu“ zu öffnen. Dann fügen Sie eine Sicherheitsausnahme für Transkribus hinzu.
- d. Nachdem Sie Transkribus gedownloadet haben, muss die Datei „entpackt“ werden. Das Programm kann nicht mit der Zip-Datei gestartet werden.

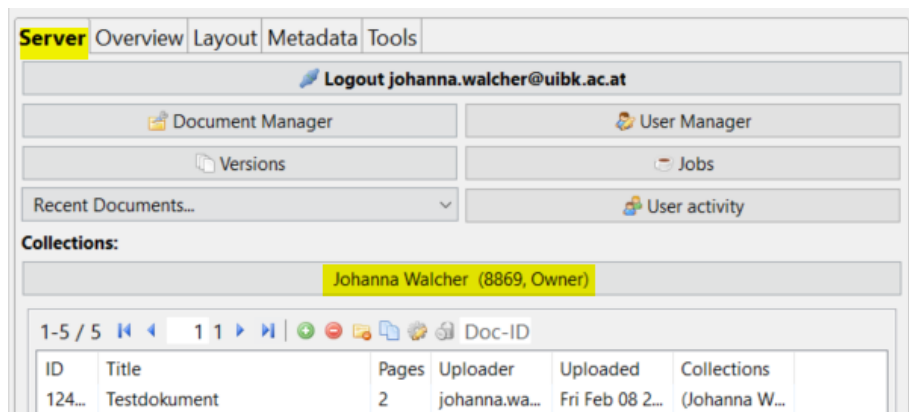
4. Transkribus öffnen

- a. Starten Sie das Programm und klicken Sie auf die „Login“ Schaltfläche im „Server“ Tab.



Darstellung 1 Login

- b. Gleich beim Start haben Sie Zugriff auf Ihre private Kollektion (benannt nach Ihrer Mailadresse). Sie beinhaltet einige Dokumente zum Experimentieren. Ihre Collections finden Sie im „Server“ Tab.



Darstellung 2 „Collections“ Schaltfläche, um zu den Testdokumenten zu gelangen

5. Dokumente hochladen

- Mit Transkribus können Sie entweder lokal an Ihren Dokumenten arbeiten, oder indem Sie sie auf den Server hochladen.
- Automatisierte Prozesse können nur ausgeführt werden, wenn Sie Ihr Dokument in die Transkribus Plattform hochgeladen haben. Die Plattform kann PDF, JPEG, PNG und TIFF Dateien verarbeiten. JP2 Dateien werden leider nicht unterstützt.
- Sie können Dokumente verwenden, die Sie selbst gescannt haben. Sie können auch unsere DocScan App für Android Smartphones verwenden, um Bilder aufzunehmen und diese dann direkt in die Transkribus Plattform hochladen. Mehr Informationen finden Sie hier: <https://scantent.cvl.tuwien.ac.at/en>
- Sie können auch Dokumente vom Internet downloaden und diese dann in die Transkribus Plattform laden. Viele Bibliotheken und Archive arbeiten mit „Open Access“ und unterstützen deshalb die Weiterverarbeitung Ihrer Kollektionen. Sie können bei Archiven und Bibliotheken direkt anfragen, ob Sie Images von Ihren Dokumenten in die Transkribus Plattform laden können.
- Klicken Sie auf die „Import document(s)“ Schaltfläche, um Bilder von Ihrem Computer auf die Plattform zu laden. Die Bilder müssen dazu in einem eigenen Ordner abgespeichert sein.



Abbildung 3 Dokumente auf die Transkribus Plattform laden

- f. Sie können die Dokumente zu einer Ihrer bestehenden Kollektionen hinzufügen oder eine neue Kollektion dafür öffnen. Das machen Sie, indem Sie auf die „Add to collection“ Schaltfläche klicken und dann auf „Create“

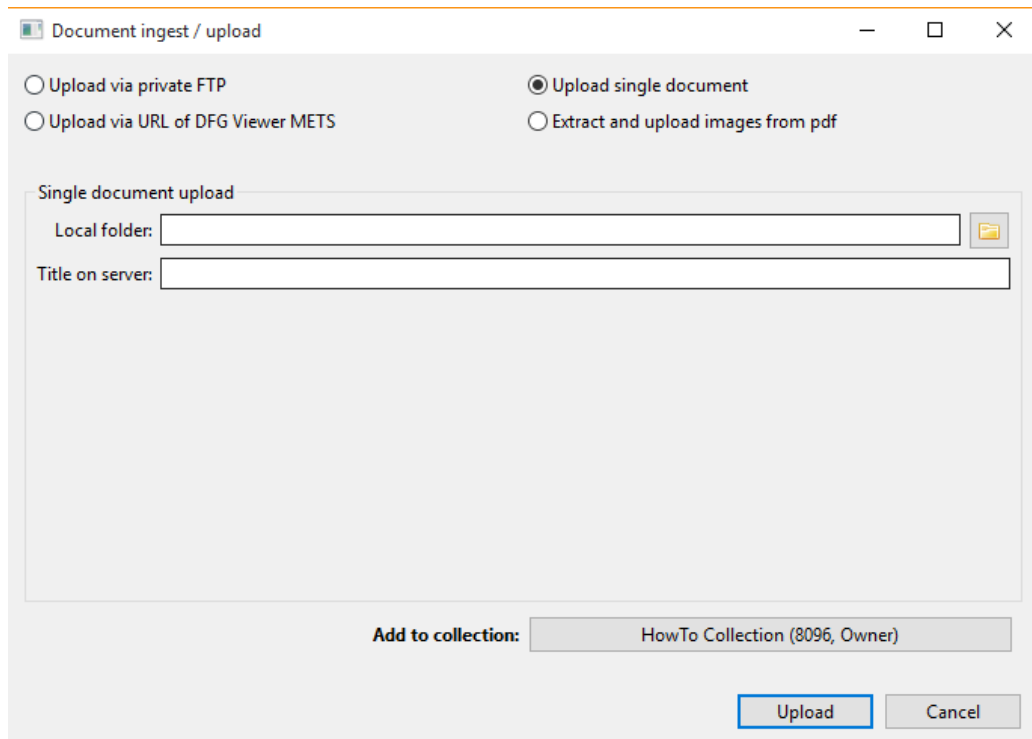


Abbildung 4 Dokumente zu einer existierenden Kollektion hinzufügen oder eine neue Kollektion öffnen

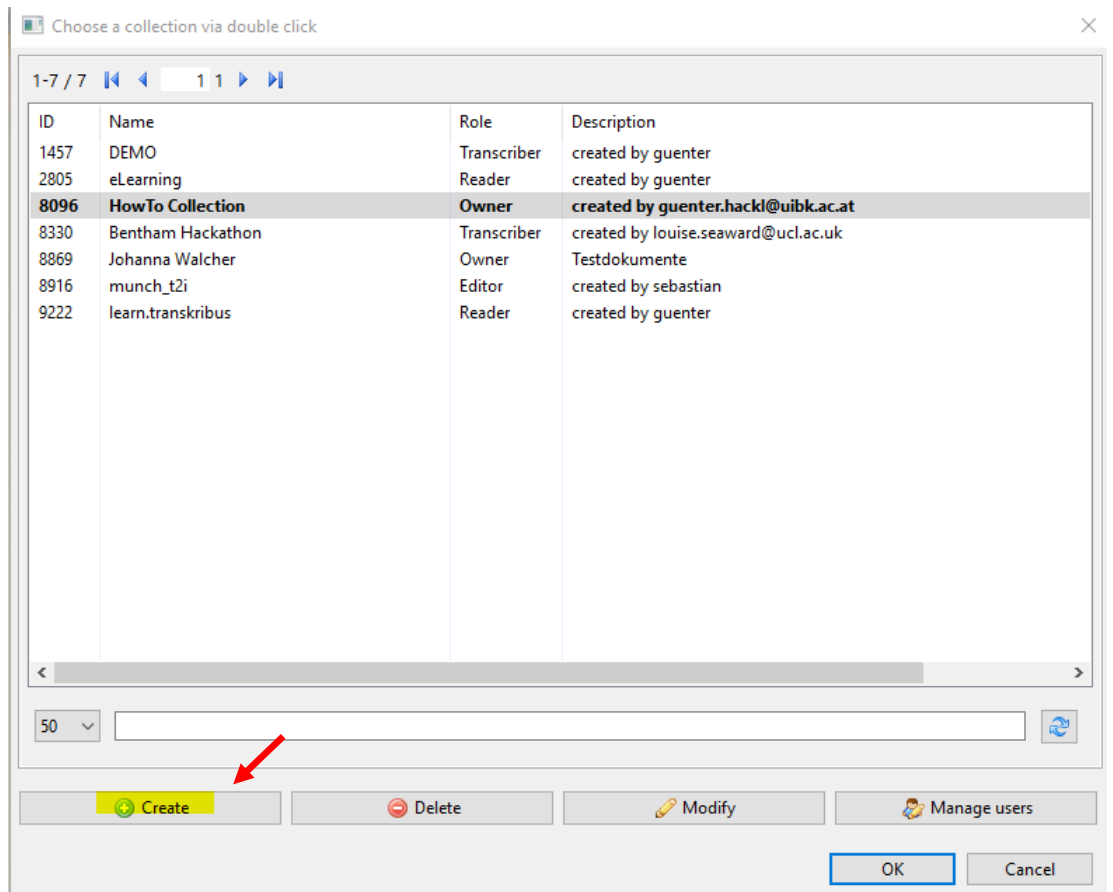


Abbildung 5 Eine neue Kollektion öffnen

- g. Um auf Ihre Dokumente zuzugreifen, klicken Sie auf die „Collections“ Schaltfläche im „Server“ Tab. Dann doppelklicken Sie auf das gewünschte Dokument, um es zu öffnen.

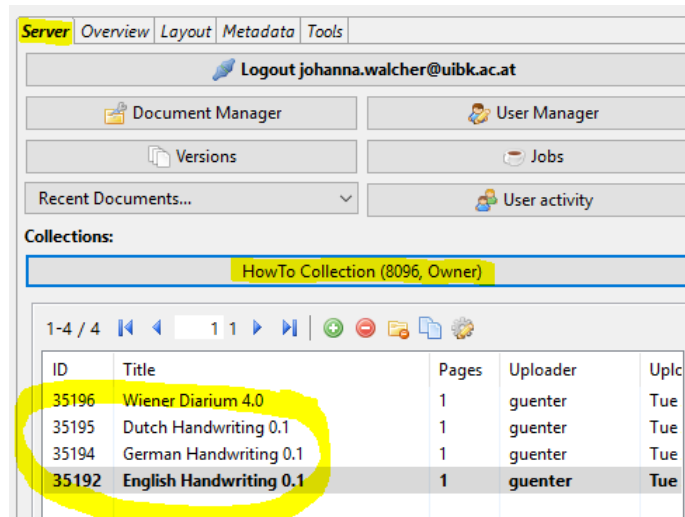


Abbildung 6 Dokumente öffnen

- h. **Alle Dokumente, die Sie in die Transkribus Plattform laden sind standardmäßig privat.** Sie können anderen Nutzern Zugriff auf die Dokumente gewähren, wenn Sie möchten. Dazu klicken Sie auf die „User Manager“ Schaltfläche im „Server“ Tab. Dokumente können nur mit Nutzern geteilt werden, die ebenfalls ein Transkribus Konto besitzen.

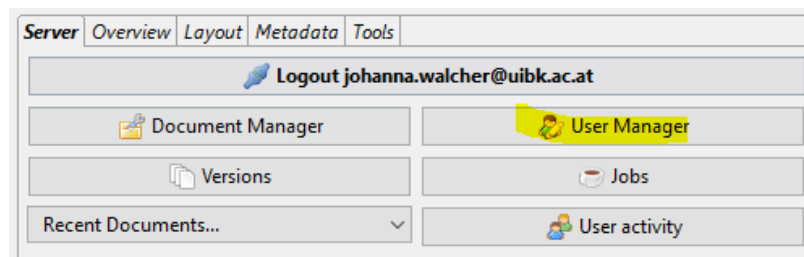


Abbildung 7 „User Manager“ Schaltfläche, um den Zugang zu Ihren Dokumenten zu verwalten

6. Dokumente segmentieren

- Um die Handschriftenerkennung mit Trainingsdaten zu füttern, müssen die Dokumente in Regionen und Linien segmentiert werden. Das geschieht automatisch.
- Sie brauchen dazu nur den „Tools“ Tab zu öffnen...
- ...und dort sicherzustellen, dass bei „Find Text Regions“ ein Häkchen gesetzt ist und dann auf „Run“ klicken.
- Sie können eine einzelne Seite segmentieren, oder das gesamte Dokument.
- Die Zeilen und Textregionen werden dann automatisch erkannt.

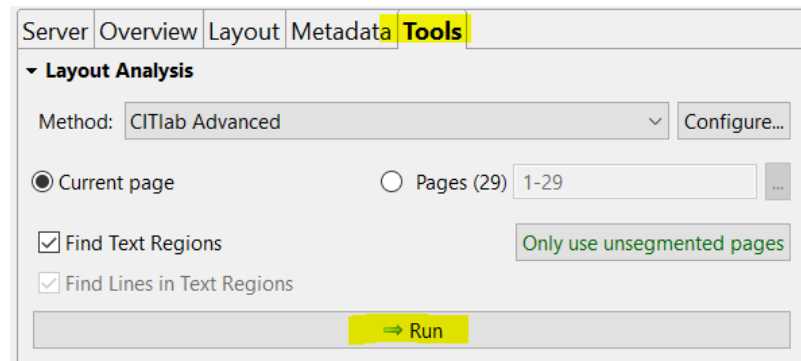


Abbildung 8 Segmentierung

7. Transkription starten

- Sobald die Baselines im Bild sichtbar sind, kann man den Transkriptionstext in das Text Editor Feld tippen.
- Klicken Sie auf die „Viewing Profiles“ Schaltfläche im Hauptmenu und wählen Sie die „Transcription“ Ansicht.
- Für jede Zeile im Bild, gibt es eine zugehörige Zeile im Text Editor. Transkribieren Sie den Text Zeile für Zeile, genauso, wie der Text im Bild erscheint.

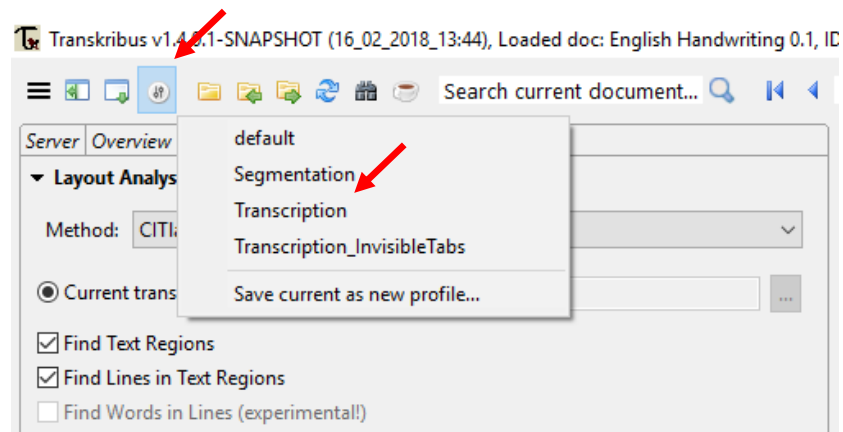


Abbildung 9 Transkriptionsansicht

- Sonderzeichen** findet man mit der „Virtual Keyboards“ Schaltfläche in der Text Editor Toolbar.



Abbildung 10 „Virtual Keyboards“ Schaltfläche

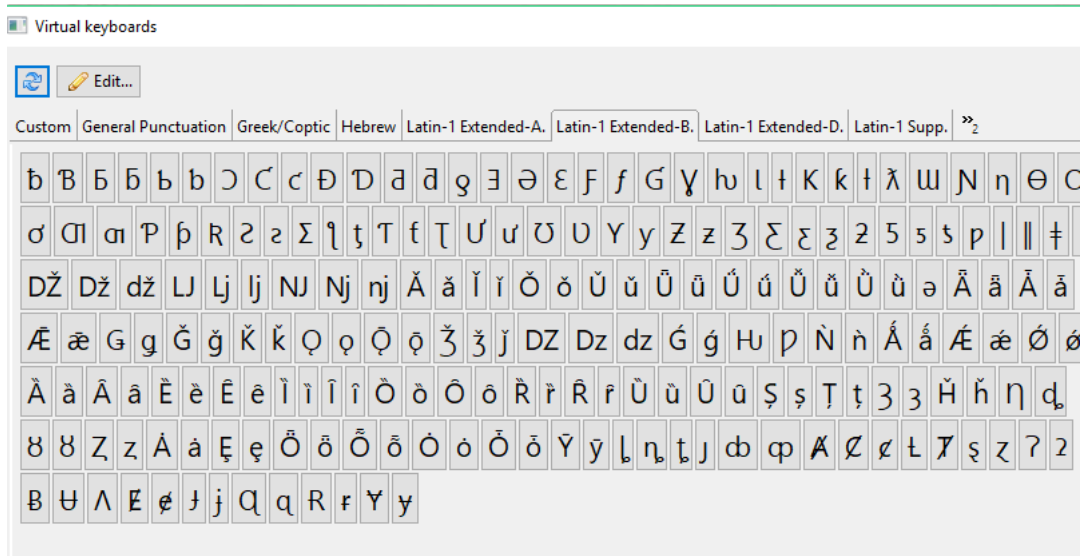


Abbildung 11 Sonderzeichen im „Virtual keyboard“

- e. Wenn Sie im Team arbeiten, kann es einfacher sein, mit der **Transkribus Web Interface** zu transkribieren. Das ist eine Lite-Version von Transkribus und sehr einfach zu nutzen: <https://transkribus.eu/read>

8. Speichern und Exportieren

- a. Klicken Sie auf die „Save“ Schaltfläche im Hauptmenu, um Ihre Dokumente in Transkribus zu speichern.

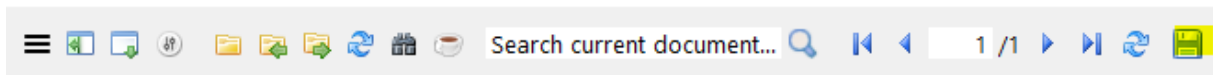


Abbildung 12 Änderungen in Ihren Dokumenten speichern

- b. Mit der „Versions“ Schaltfläche im „Server“ Tab sehen Sie, dass eine **neue Version** des Dokuments erstellt wurde. Sie können hier immer wieder zu den vorhergegangenen Versionen zurückkehren (durch Doppelklick) und haben so kein Risiko, Daten bzw. getane Arbeit zu verlieren.

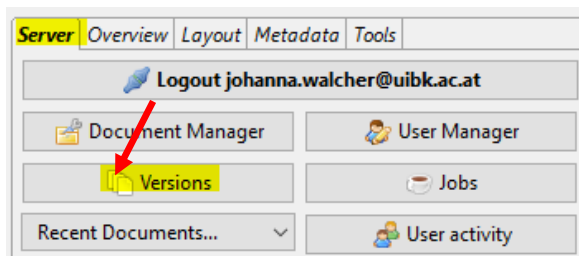


Abbildung 13 „Versions“ Schaltfläche für Zugang auf vorhergegangene Versionen des Dokuments

- c. Außerdem können Sie das **gesamte Dokument** jederzeit und unabhängig davon, wie weit Sie mit der Arbeit an dem Dokument schon sind, **exportieren**. Dazu klicken Sie auf die „Export document“ Schaltfläche im Hauptmenu.



Abbildung 14 „Export document“ Schaltfläche

9. Handschriftenerkennung (Handwritten Text Recognition – HTR+) auf Ihre Dokumente anwenden

- a. Ganz einfach können Sie Ihre Dokumente vom Computer erkennen lassen. Mit ca. 5 000 Wörter von gedrucktem Text oder 15 000 Wörter von handgeschriebenem Text können Sie das Training eines Modells starten. Um den Trainingsprozess zu starten, schicken Sie uns bitte eine kurze Mail, sobald Sie die Dokumente segmentiert und schon einige Seiten transkribiert haben: email@transkribus.eu.
- b. Sie bekommen dann von uns die Berechtigung für das Modell-Training in Transkribus. Für weiter Informationen zum Modell-Training, konsultieren Sie bitte folgende Anleitung: [Modell Training in Transkribus](#)
- c. Sobald das Training des Modells abgeschlossen ist, können Sie es an Ihren Dokumenten anwenden. Öffnen Sie dazu den „Tools“ tab und klicken Sie auf die „Run“ Schaltfläche im „Text Recognition“ Abschnitt. Sie können eine oder mehrere Seiten für die Erkennung auswählen.

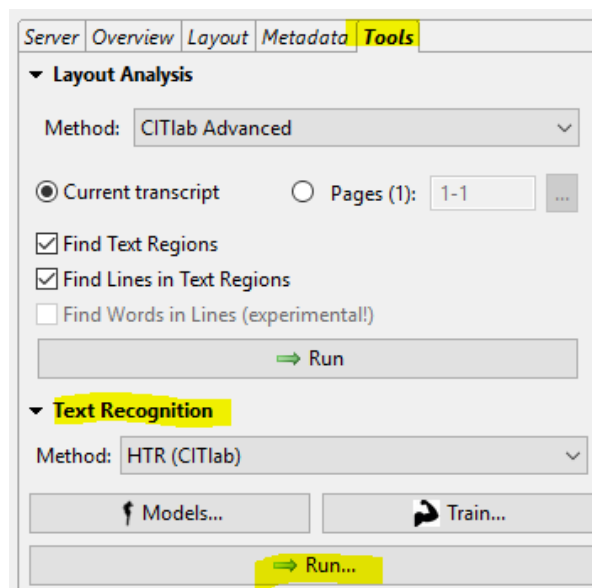


Abbildung 15 Handschriftenerkennung anwenden

The screenshot shows the 'Choose a model' dialog box. On the left, a table lists various models with columns for Name, Language, and Curator. The 'English Writing M1' model is selected. On the right, the 'Details' section provides information about the selected model, including its name, language, description, and a learning curve graph. The graph shows 'Accuracy in CER' on the y-axis (0% to 100%) and 'Epochs' on the x-axis (0 to 80). Two lines are plotted: a blue line for 'CER Train' and a red line for 'CER Test'. The CER on Train Set is 4.54% and the CER on Test Set is 3.66%.

Name	Language	Curator
English Writing M1	English	Unkno
OEAW M3	German	guente
MS A2654	Arabic	guente
Binder Kochbuch M2	German	guente
Itinera_Nova_M1	dutch	guente
PROB 11 1840s M1 (25_09_2017)	english	guente
Thun Missiven M3	German	guente
RA_GEO_M1	English	guente
Wydean	Latin	guente
Wiener Diarium M3	German	guente
GNM_typoskript	german	guente
Konzilsprotokolle v1	German	philip
Konzilsprotokolle M4	german	guente
ReiserDrosteWrDiariumWurzbach	German	guente
HHStA-KK M2	German	guente
Hebrew_t2i_test	hebrew	guente
Liber Extended M8	Latin	guente
StazH_Protokolle_t2i	german	guente
German Kurrent (Reichsgericht)	German	guente
Itinera-Thun-M1	dutch	guente

Details

Name: English Writing M1 Language: English

Description: Based on Jeremy Bentham and secretaries from early 19th century. 50.000+ words. Parameters: N/A N/A

Nr. of Words: N/A Nr. of Lines: N/A

Buttons: Show Train Set Show Test Set Show Character Set

Learning Curve

Accuracy in CER vs Epochs

Legend: CER Train (blue line), CER Test (red line)

CER on Train Set: 4.54% CER on Test Set: 3.66%

Abbildung 16 Überblick über die Modelle und Lernkurve

- Wenn Sie auf „Run“ und dann auf „Configure“ klicken, sehen Sie Informationen über das Modell.
- Auf der linken Seite des Fensters sehen Sie einen Überblick über die verfügbaren Modelle.
- Oben rechts sehen Sie die Details zum Modell.
- Die Grafik unten rechts zeigt die Genauigkeit des Modells in „Character Error Rate (CER)“, also der Prozentsatz von Zeichen, der von der Handschriftenerkennung falsch transkribiert wurde. Die blaue Linie zeigt den Fortschritt des Trainings. Die rote Linie zeigt den Fortschritt in der Evaluation mit dem „Test Set“. Das „Test Set“ sind jene Seiten des Dokuments, die im Training ausgespart wurden, um im Nachhinein die Effektivität des Modells zu testen.
- Nachdem die Handschriftenerkennung abgeschlossen ist, erscheinen die Ergebnisse in einer neuen Version des Dokuments. Mit der „Compute Accuracy“-Funktion im „Tools“ Tab kann die automatisch generierte Transkription evaluiert werden, indem man Sie mit der „Ground Truth“ Version des Dokuments vergleicht.

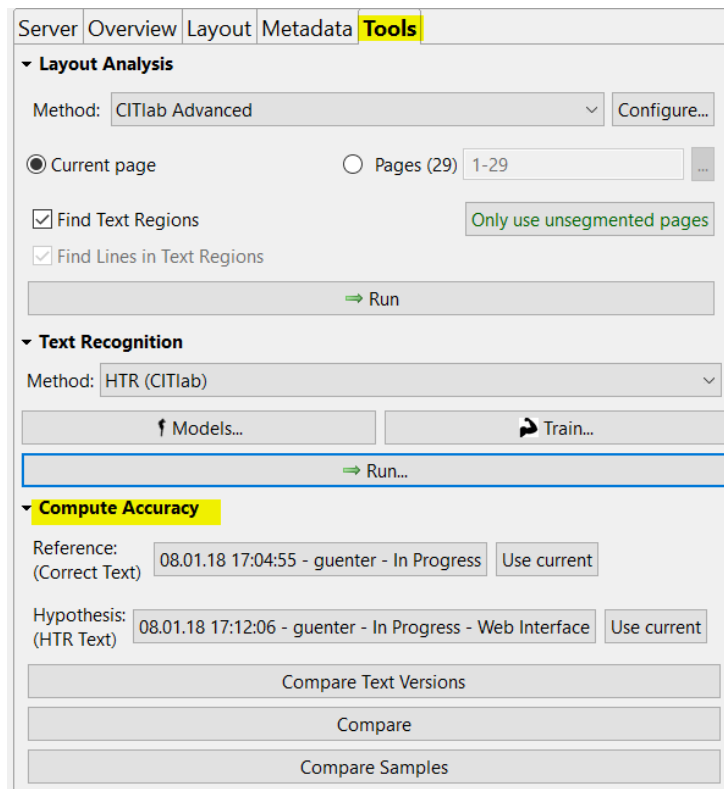


Abbildung 17 Genauigkeit der Handschriftenerkennung berechnen

10. Keyword Spotting

- Mit einem Handschriftenerkennungsmodell können Sie in Ihren Dokumenten mit der Keyword Spotting Funktion nach bestimmten Wörtern suchen.
- Lassen Sie ein HTR-Modell über Ihre Dokumente laufen, um eine automatische Transkription zu generieren.
- Dann öffnen Sie das Keyword Spotting, indem Sie auf die Fernglas-Schaltfläche im Hauptmenu klicken (siehe Abbildung 18).



Abbildung 18 Das "Search for..." Fenster öffnen um das Keyword Spotting zu nutzen

- Wählen Sie den „KWS“ Tab in dem sich öffnenden Fenster.

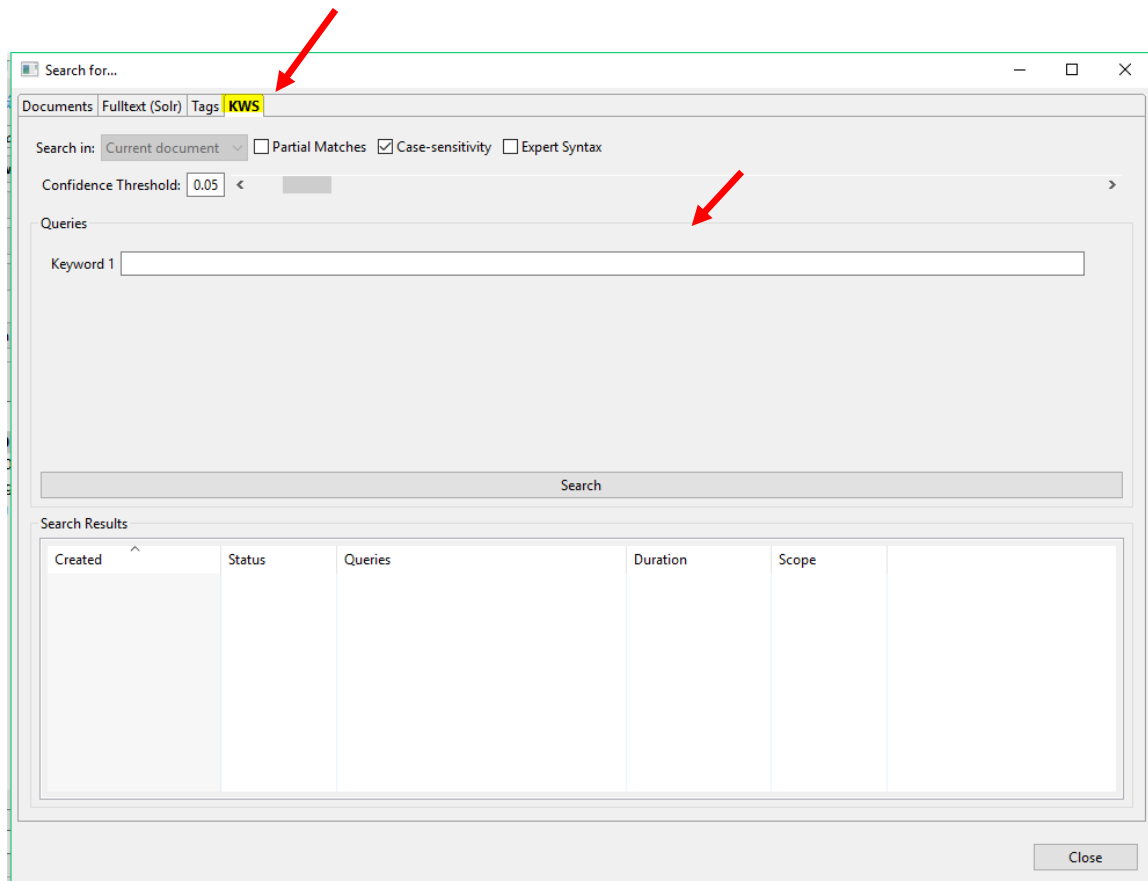


Abbildung 19 „Search for...“ Fenster

- e. Tippen Sie das Wort, nachdem Sie gerne suchen möchten in das „Keyword 1“ Feld und klicken Sie auf „Search“.
- f. Dann erscheint ein Bestätigungsfenster. Klicken Sie auf „Yes“, um die Suchanfrage zu starten.

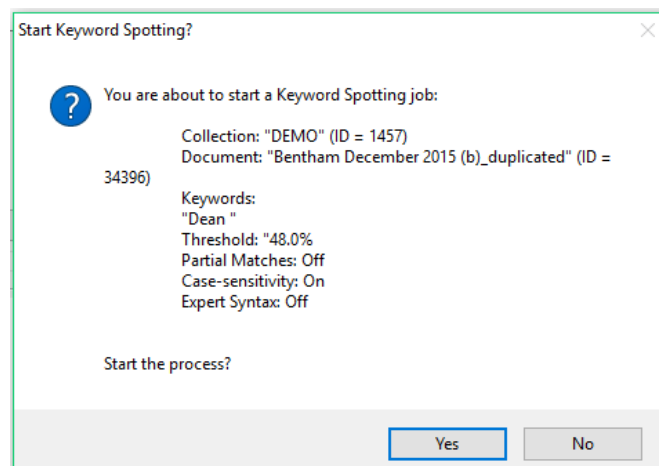


Abbildung 20 Bestätigungsfenster

11. Suchergebnisse des Keyword Spotting

- a. Sobald die Suchanfrage erledigt ist, doppelklicken Sie auf das Datum in der „Created“ Spalte, um zu den Suchergebnissen zu gelangen.

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...

Abbildung 21 Keyword Spotting Ergebnisse

- b. Das „Keyword Spotting Results“ Fenster zeigt eine Liste mit Textausschnitten, in denen das Suchwort gefunden wurde.

Keyword Spotting Results

"Dean " (10 hits)

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

Abbildung 22 Informationen über die Keyword Spotting Suchergebnisse

Danksagung

Wir möchten den vielen Nutzern danken, die mit ihrem Feedback zur Verbesserung der Transkribussoftware beigetragen haben.