

READ

Recognition and Enrichment
of Archival Documents



Come usare Transkribus – in 10 passi (o anche meno)

*Version v1.4.0. (22_02_2018_15:07)
Last update of this guide 04.04.2018*

Questo documento è un'introduzione di base a Transkribus. Il testo fornisce le informazioni per svolgere un flusso di lavoro standard usando la piattaforma. Se hai bisogno di informazioni più dettagliate sulle funzioni di Transkribus puoi visitare la nostra guida in formato Wiki: <https://transkribus.eu/wiki/>

Scarica Transkribus Expert Client, o assicurati di usare la versione più aggiornata:

- <https://transkribus.eu/>

Leggi la guida Transkribus Wiki per avere ulteriori dettagli e altre guide How To:

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

Il progetto Transkribus e la tecnologia con la quale è stato sviluppato sono disponibili nei seguenti siti:

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

Contatti

- Il team Transkribus: email@transkribus.eu



Il progetto READ ha ricevuto finanziamenti dal programma di ricerca e innovazione dell'Unione Europea Horizon 2020 con la convenzione numero 674943.

1. Introduzione

- a. Transkribus può essere usato per diversi scopi. I più importanti sono:
 - a.i. Trascrivere documenti per un'edizione critica
 - a.ii. Creare un set di dati per effettuare il training del sistema **Handwritten Text Recognition (HTR)** affinché possa imparare a decifrare i tuoi documenti storici.
 - a.iii. Eseguire HTR sui tuoi documenti e ricevere **automaticamente la trascrizione**
 - a.iv. Cercare parole distinte nelle tue raccolte di documenti utilizzando **Keyword Spotting** decisamente più performante della classica ricerca full-text.
 - a.v. La piattaforma vive grazie alla comunità. Più dati vengono caricati su Transkribus e più efficiente sarà il programma, specialmente il sistema HTR (riconoscimento del testo scritto a mano).
- b. Transkribus è offerto come una infrastruttura di ricerca grazie al progetto H2020 **READ (Recognition and Enrichment of Archival Documents)** <https://read.transkribus.eu/>.
- c. Prenditi del tempo per esplorare Transkribus e acquisire familiarità con il suo funzionamento. Per rendere tutto più semplice abbiamo creato molte guide come questa che forniscono istruzioni sulle diverse funzioni della piattaforma. Puoi trovarli nella Wiki di Transkribus: <https://transkribus.eu/wiki/>

2. Registrati sul sito

- a. Vai su: <http://transkribus.eu/>
- b. Leggi la nostra **licenza d'uso**: https://transkribus.eu/Transkribus/docs/TranskribusTermsOfUse_v04-2016.pdf
- c. Tutti i documenti caricati su Transkribus sono "privati", questo vuol dire che nessuno, a parte te, potrà accedervi.
- d. Il team Transkribus rispetta tutte le direttive europee in termini di protezione dati e privacy. Rispetteremo la tua privacy e useremo i dati solo per migliorare i nostri servizi e sostenere la ricerca nelle discipline umanistiche e informatiche!

3. Scarica Transkribus dal sito web

- a. Vai sul sito Transkribus <http://transkribus.eu/> e clicca su "Download".
- b. Transkribus funziona su Windows, MacOS and Linux. Se hai bisogno di aiuto per installare la piattaforma puoi consultare la Wiki: https://transkribus.eu/wiki/index.php/Download_and_Installation
- c. Se usi MacOS potrebbe apparire un messaggio di errore la prima volta che provi ad aprire Transkribus. In questo caso:
 - c.i. Clicca con il pulsante destro del Track Pad per aprire il Menu Contestuale e aggiungere una eccezione di sicurezza per Transkribus.

- d. Una volta scaricato Transkribus, scompatta il file. Il programma non può partire da un file zippato!

4. Apri Transkribus

- a. Avvia il tool e clicca il pulsante “Login” nella scheda “Server”.

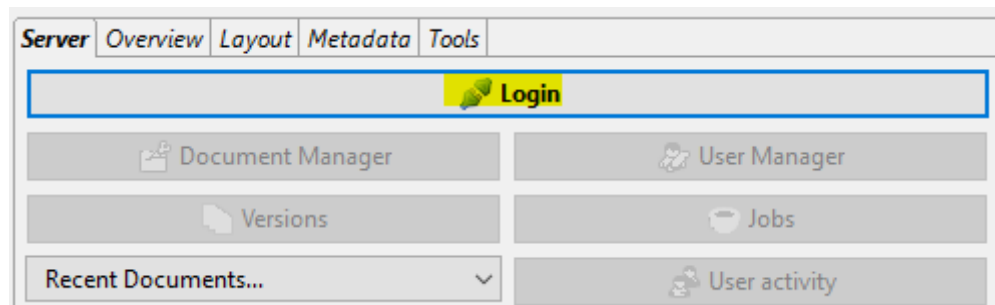


Figura 1 Login

- b. Avrai accesso alla tua collezione privata che prende il nome dal tuo indirizzo email. Questa raccolta include alcuni documenti di prova che puoi usare per fare delle prove.
- c. Puoi trovarli cliccando il pulsante “Collections” nella scheda “Server”.

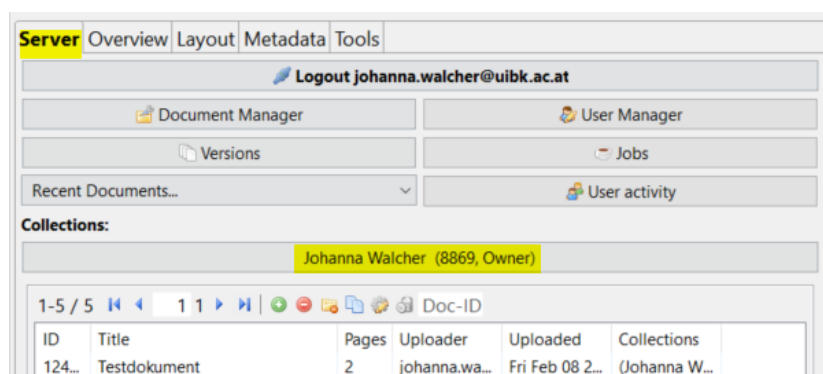


Figura 2 Testa i documenti nella tua collezione

5. Carica I tuoi documenti

- a. Transkribus ti permette di lavorare con i tuoi documenti, sia localmente che caricandoli sul server.
- b. I processi automatici possono essere eseguiti solo se caricati sulla piattaforma Transkribus. La piattaforma può processare PDF, JPEG, PNG e file TIFF. Non supporta il formato JPEG 2000 sfortunatamente.
- c. Puoi caricare i documenti che hai scansionato tu stesso. Puoi usare la nostra app **DocScan** per smartphone Android per catturare le immagini e caricarle direttamente su Transkribus. Per altre informazioni consulta: <https://scantent.cvl.tuwien.ac.at/en>
- d. Puoi inoltre scaricare documenti da Internet e caricarli su Transkribus. Molte biblioteche e archivi seguono la filosofia dell'Open Access, ragion per cui

incoraggiano l'utilizzo delle loro collezioni. Puoi chiedere direttamente ai responsabili degli archivi se puoi caricare le immagini dei loro documenti su Transkribus!

- e. Clicca il pulsante "Import document(s)" per trasferire le immagini dal tuo computer alla piattaforma. Nota: le immagini devono trovarsi in una cartella separata prima di poterle caricare su Transkribus!

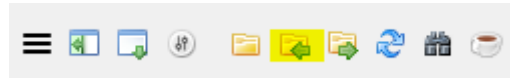


Figura 3 Carica i tuoi documenti su Transkribus

- f. Puoi aggiungere i tuoi documenti ad una delle tue collezioni già esistenti o crearne una nuova cliccando sul pulsante "Add to collection" presente alla fine della schermata "Document ingest/upload" e poi cliccando su "Create".

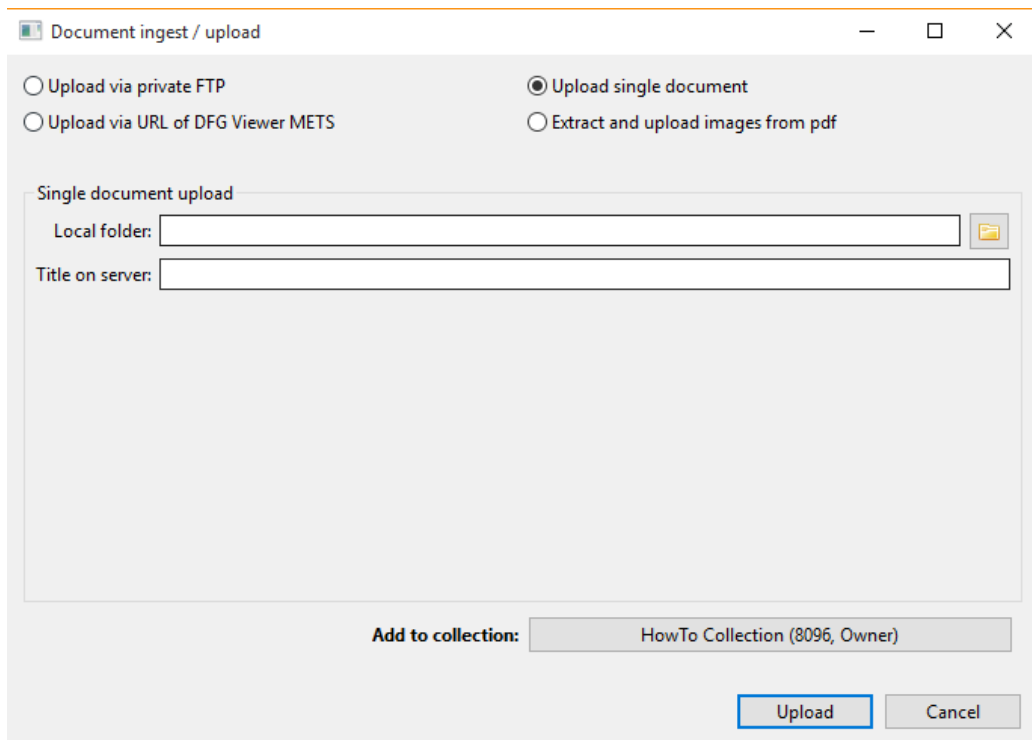


Figura 4 Aggiungi documenti ad una collezione esistente o creane una nuova

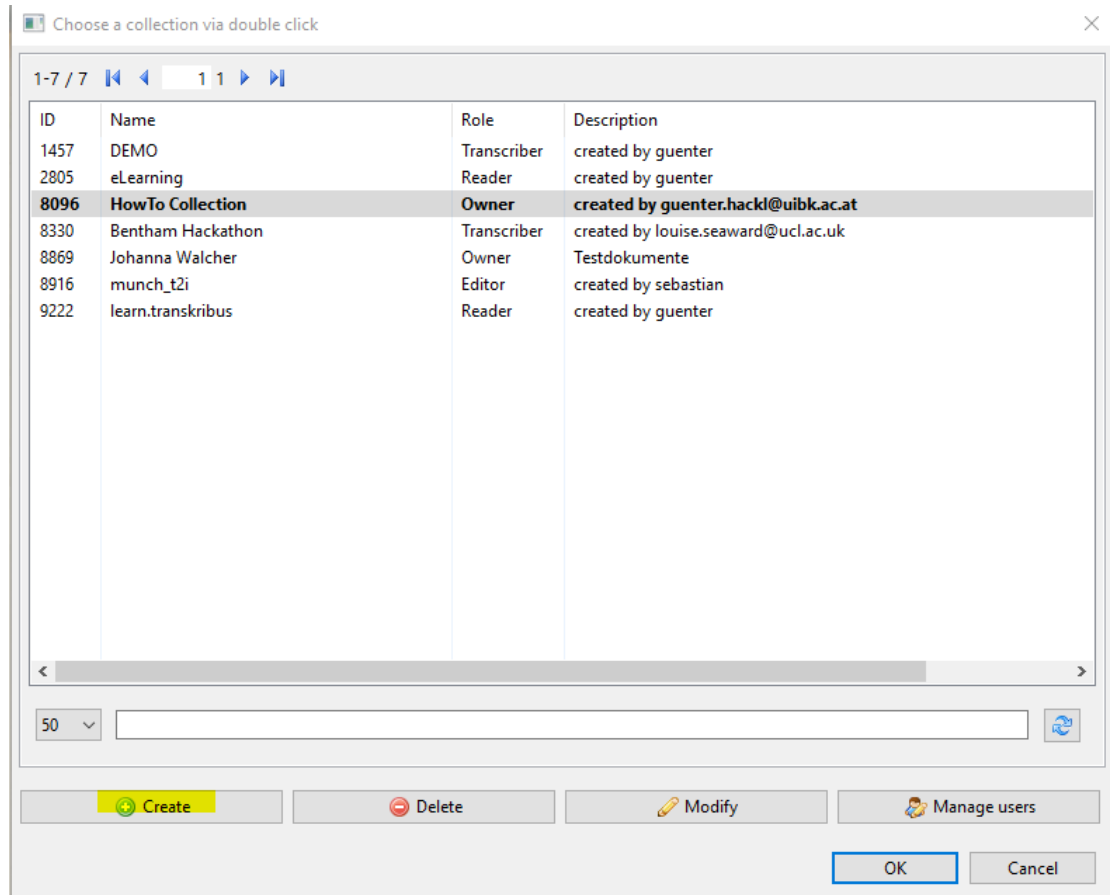


Figura 5 Crea la tua collezione

- g. Per avere accesso ai tuoi documenti, clicca sul pulsante “Collections” nella scheda “Server” e scegli la collezione. Dopo esegui un doppio clic su uno dei documenti nella schermata in basso, nella scheda “Server”, per aprirne uno.

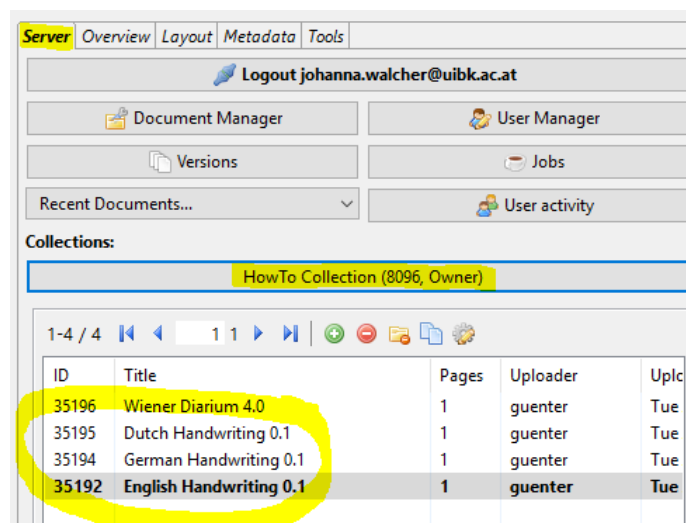


Figura 6 Apri i documenti nella tua collezione

- h. Tutti i documenti caricati su Transkribus sono privati di default.** Puoi dare ad altri utenti l'autorizzazione per vedere i tuoi documenti se vuoi. Usa il pulsante "User Manager" nella scheda "Server" per aggiungere altri utenti alle tue collezioni. Puoi condividere solo le collezioni con altri utenti che hanno un account su Transkribus.

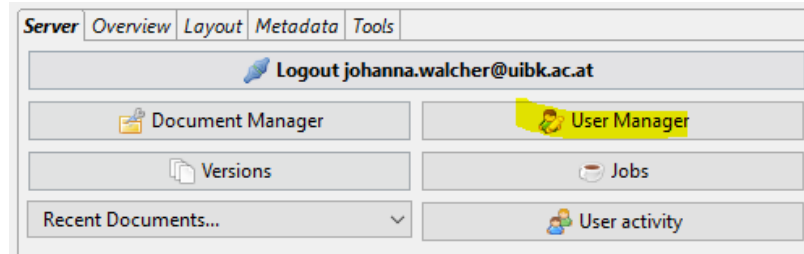


Figura 7 Usa il pulsante "User Manager" per gestire l'accesso alle tue collezioni

6. Segmenta i tuoi documenti in righe

- Per poter addestrare il motore HTR con i dati di training, i documenti devono essere segmentati in righe. Questa operazione è possibile farla in maniera automatica con Transkribus.
- Apri la scheda "Tools".
- Assicurati che "Find Text Regions" e "Find Lines in Text Regions" siano selezionati e clicca sul pulsante "Run".
- Puoi scegliere di segmentare la pagina corrente o una serie di pagine.
- Le righe e le aree di testo verranno rilevate automaticamente.

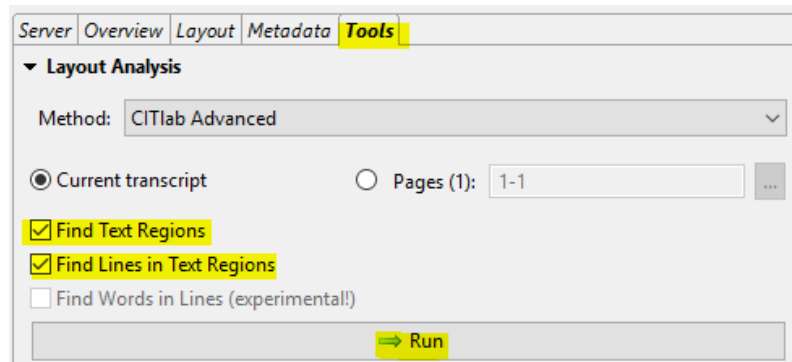


Figura 8 Segmentazione

7. Inizia la tua trascrizione

- Quando le linee di base sono visibili sull'immagine, puoi cominciare a scrivere nel campo Editor di testo (Text Editor field).
- Clicca sul pulsante "Viewing Profiles" e seleziona la vista "Transcription".

- c. Per ogni linea di base ci sarà **una riga corrispondente nell'Editor di Testo**. Trascrivi il testo riga per riga, esattamente come appare nell'immagine.

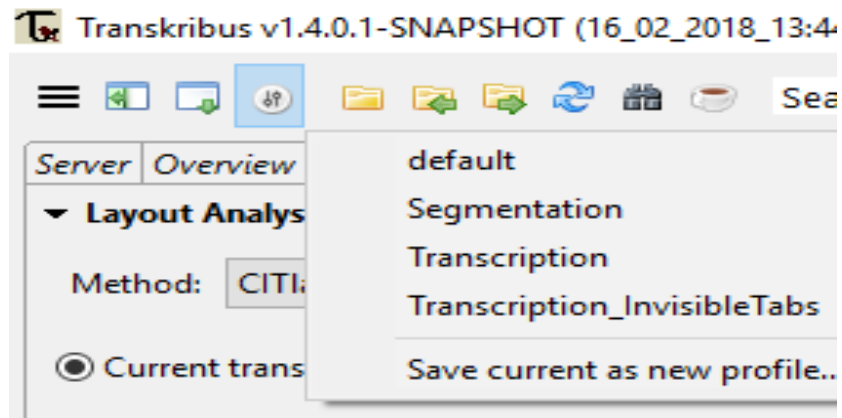


Figura 9 Transcription view

- d. I caratteri speciali si possono trovare cliccando sul pulsante **“Virtual Keyboards”** (Tastiere Virtuali) nella barra degli strumenti nell'editor di testo.

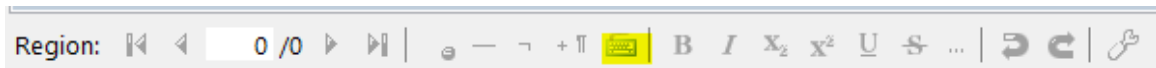


Figura 10 pulsante “Virtual Keyboards” (Tastiere Virtuali)

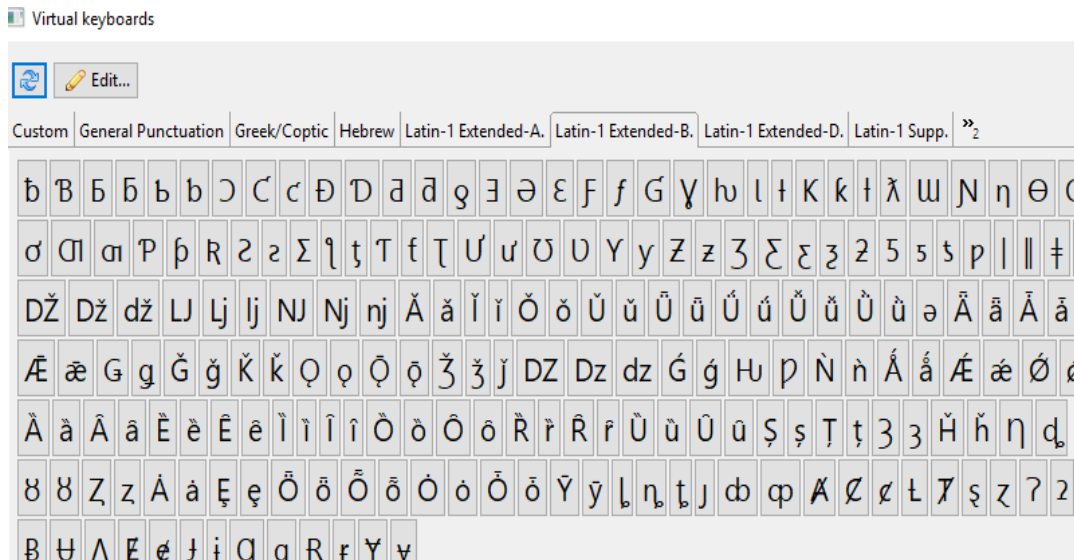


Figura 11 Virtual keyboards (Tastiere Virtuali)

- e. Se lavori in gruppo, potresti trovare più semplice trascrivere utilizzando l'**interfaccia Web di Transkribus (Transkribus Web Interface)**. Questa è una versione ridotta di Transkribus più semplice da usare: <https://transkribus.eu/read>

8. Salva ed esporta la tua trascrizione



- a. Premi il pulsante “Save” nel Menu Principale per salvare i documenti in Transkribus.

Figura 12 Salva i cambiamenti nel tuo documento

- b. Se clicchi sul pulsante “Versions” nella scheda “Server”, vedrai che una **nuova versione** è stata creata. Questo vuol dire che puoi sempre accedere alla versioni precedenti di un documento salvato se necessario.

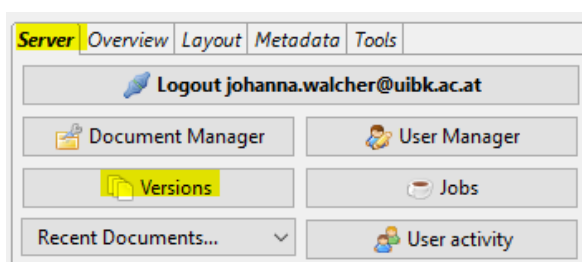


Figura 13 Clicca sul pulsante “Versions” per accedere alle versioni precedenti del tuo documento

- c. Puoi anche **esportare l'intero documento** in qualsiasi momento del processo cliccando sul pulsante “Export document”.



Figura 14 Pulsante “Export document”

9. Utilizza il riconoscimento del testo scritto a mano (HTR) nei tuoi documenti

- a. È semplice utilizzare il computer per farti riconoscere i tuoi documenti. Puoi iniziare ad effettuare il training di un modello utilizzando circa 5000 parole trascritte di testo stampato o 15000 parole di testo scritto a mano. Per iniziare il processo di training inviaci una breve email dopo aver segmentato e trascritto un primo lotto di pagine (email@transkribus.eu).
- b. Riceverai il permesso di eseguire il training del tuo modello direttamente da noi. Se hai bisogno di ulteriori informazioni puoi consultare la guida: [How to Train a Model](#).
- c. Una volta che il **modello HTR** è stato addestrato per i tuoi documenti, può essere applicato cliccando sul pulsante “Run” nella scheda “Text Recognition” nella scheda “Tools”. Puoi selezionare una o più pagine dei tuoi documenti e avviare il riconoscimento automatico.

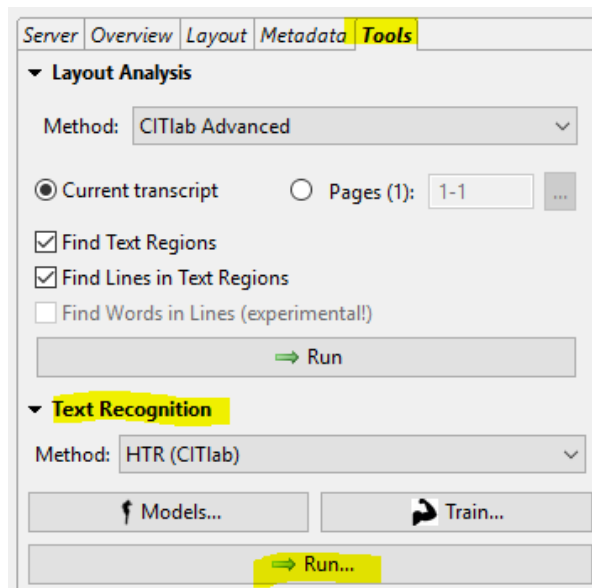


Figura 15 Avvia il riconoscimento della scrittura a mano (HTR)

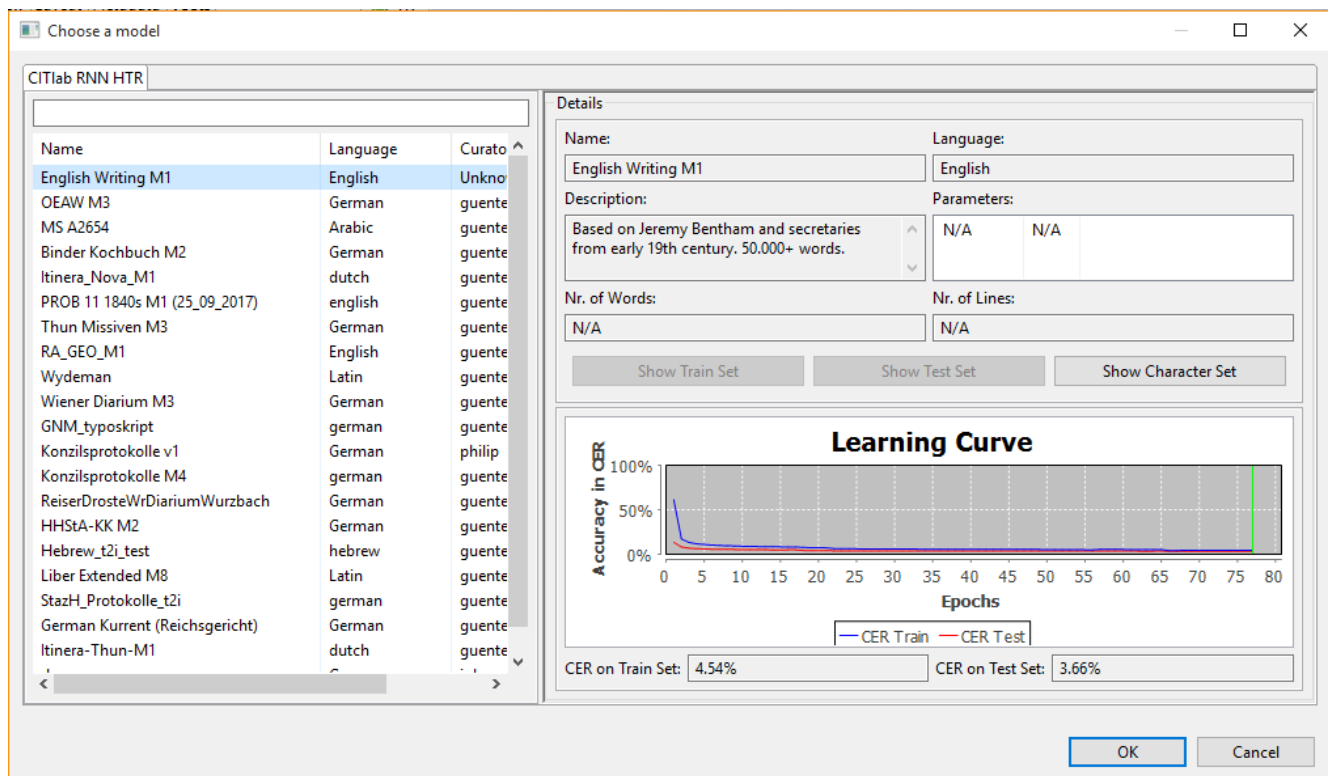


Figura 16 Panoramica del modello e curva di apprendimento

- d. Se clicchi su “Run” e poi su “Configure”, vedrai le informazioni sul tuo modello.
- e. Sul lato sinistro della finestra puoi vedere una panoramica dei modelli disponibili.
- f. Sul lato destro della finestra sono mostrati i dettagli del modello.
- g. Il grafico in basso a destra indica la precisione del modello con il Tasso di Errore del Carattere (CER - Character Error Rate), ovvero la percentuale di caratteri che sono stati trascritti in modo errato dall'HTR. La **linea blu** rappresenta il progresso del training. La linea rossa rappresenta lo stato di avanzamento delle valutazioni sul set di test dei dati che è stato messo da parte durante il processo di addestramento.

- h. Dopo che il riconoscimento con HTR è stato effettuato, i risultati verranno visualizzati direttamente su una nuova versione del documento in Transkribus. È possibile **valutare l'accuratezza** della trascrizione automatica utilizzando il pulsante “Compare” nella sezione “Compute Accuracy” nella scheda “Tools”.

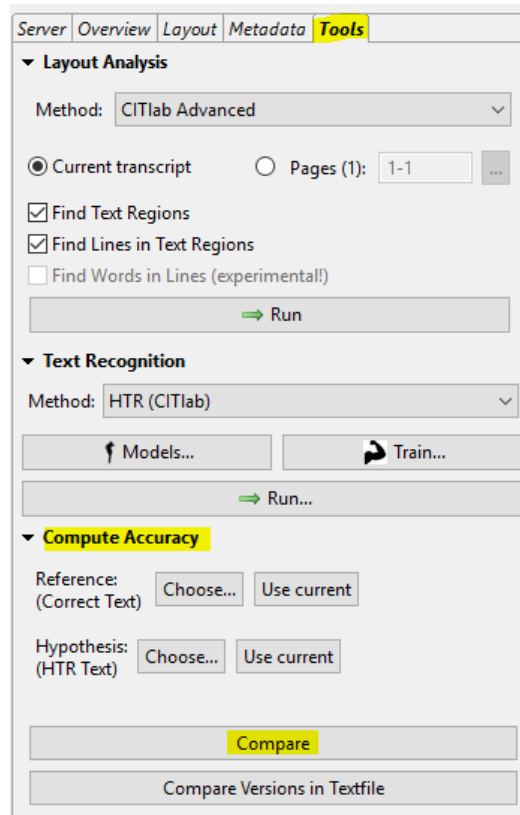


Figura 17 Calcola la precisione dell'HTR

10. Keyword Spotting

- a. Adesso che hai un modello HTR per i tuoi documenti, puoi effettuare ricerche utilizzando la funzione Keyword Spotting
- a. Prima di tutto avvia un modello HTR sui tuoi documenti per creare una trascrizione automatica
- b. La funzione Keyword Spotting sarà disponibile cliccando sul pulsante con il binocolo come mostrato in Figura 18.



Figura 18 Apri la finestra “Search for...” per usare la funzione Keyword Spotting

c. Nella finestra apri e scegli la scheda “KWS”.

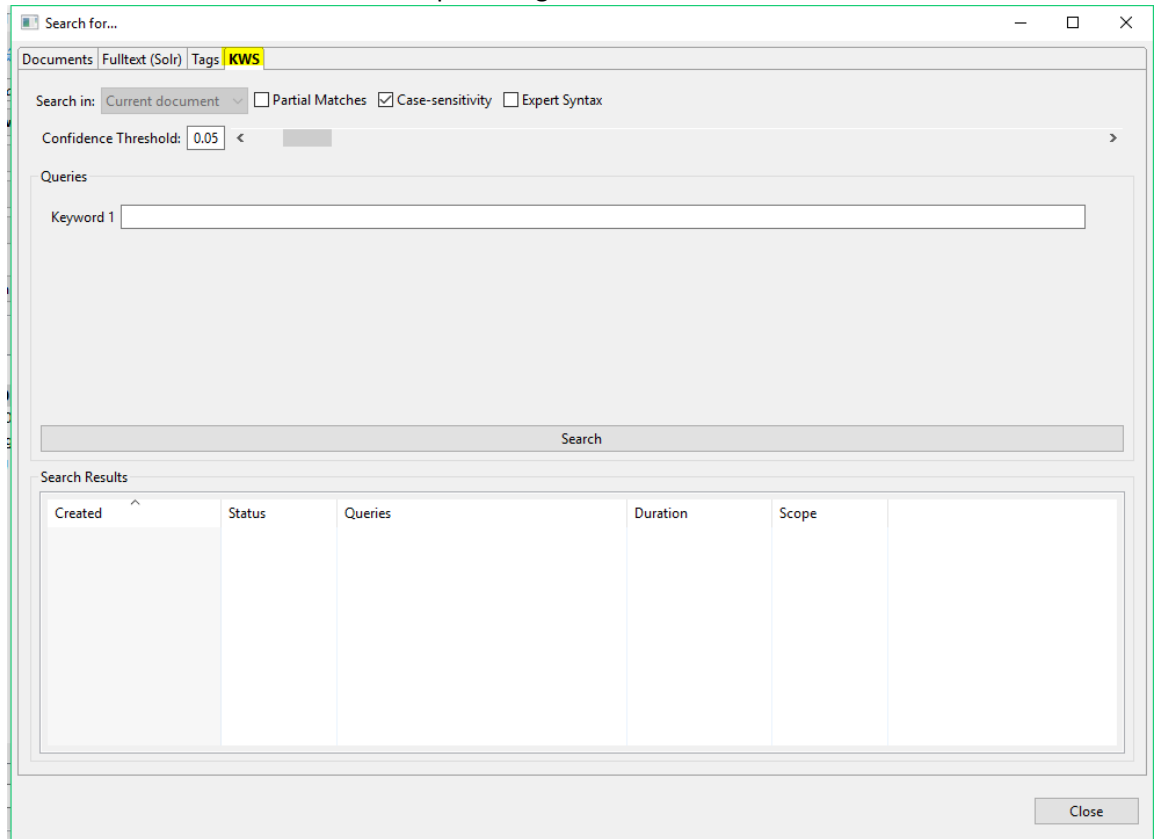


Figura 19 Finestra per utilizzare la funzione Keyword Spotting

- Digita semplicemente la parola che vuoi cercare nella casella “Keyword 1” e premi il pulsante “Search”.
- Una finestra di conferma verrà visualizzata. Clicca “Yes” per avviare l'interrogazione utilizzando la funzione Keyword Spotting

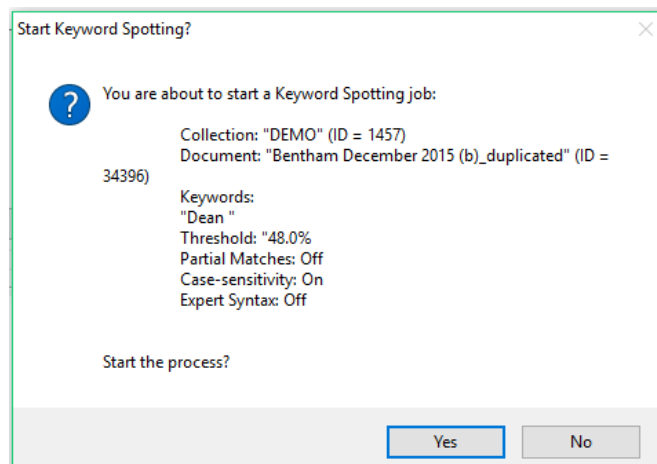


Figura 20 Finestra di conferma

- Una volta terminata la fase di ricerca, fai doppio clic sulla data e sul valore numerico nella colonna “Created” per poter accedere ai risultati della tua interrogazione.

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 21 Risultati Keyword Spotting

- La finestra "Keyword Spotting Results" ti mostrerà un elenco dei posti in cui viene visualizzata la parola chiave

The screenshot shows a window titled "Keyword Spotting Results" with a sub-header "Dean " (10 hits). The main content is a table with the following columns: Confidence, Page Nr., Line transcription, and Previ... (Preview). The table lists 10 results with varying confidence scores and page numbers. The transcription for the top result is "transmitted to his Lordship by the Dean; which". To the right of the transcription column, there are small thumbnail images of the original document pages. Below the table, there is a "Preview" section showing a larger view of the document page with the word "Dean" clearly visible in cursive handwriting.

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	

Figura 22 Informazioni sui risultati relativi alla funzione Keyword Spotting

Ringraziamenti

Vogliamo ringraziare i molti utenti che hanno contribuito con il loro feedback per aiutare e migliorare il software Transkribus.

Transkribus è reso disponibile al pubblico come parte del progetto H2020 e-Infrastructure Project READ (Recognition and Enrichment of Archival Documents) che ha ricevuto finanziamenti dall'Unione Europea.