PRIMER UPORABE PROGRAMA TRANSKRIBUS IN IZDELAVA MODELA ZA AVTOMATSKO OPTIČNO PREPOZNAVANJE ZNAKOV ZA POENOSTAVLJENO TRANSKRIBIRANJE PISANE GOTICE

DOSTOPNOST ARHIVSKEGA GRADIVA S STRANI UPORABNIKOV

Izzivi:

- fizična omejitev (le en uporabnik na arhivsko škatlo)
- nerazumevanje gradiva zaradi oddaljenosti obdobja nastanka (dokument zapisan v tujem jeziku, tuji pisavi (!)

Rešitev: Prilagoditev in predstavitev arhivskega gradiva za potrebe uporabnikov

Konkretni primer:

Arhivski dokumenti v ročno pisani gotici niso razumljivi vsem uporabnikom zato je potreben proces transkribiranja.

KAJ JE TRANSKRIBUS

Programsko orodje, katerega namen je narediti arhivsko gradivo bolj dostopno.

Razvito s strani Innsbruške univerze v okviru evropskega projekta READ.

Namen orodja je kot pomoč pri transkripciji ročno pisanih dokumentov.

Omogoča avtomatizirano prepoznavanje ročno zapisanih dokumentov.

Namen uporabe povsod, kjer so hranjeni zapisi dokumentov (se lahko v knjižnicah, muzejih, arhivih).

https://transkribus.eu/Transkribus/

KAKO DELUJE TRANSKRIBUS

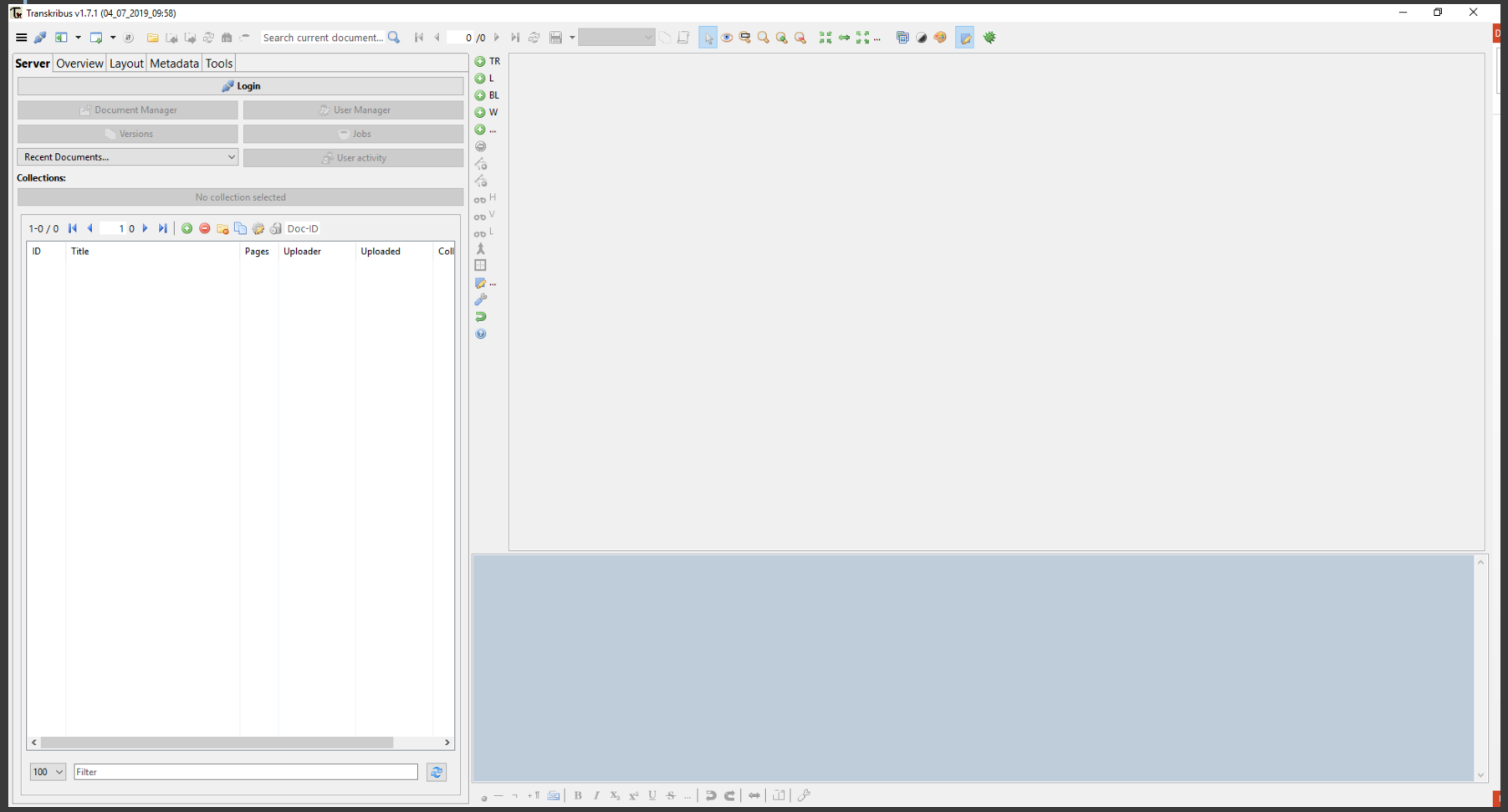
Sam program omogoča, da digitalizate arhivskega gradiva, katere želimo obdelovati, uvozimo v programski vmesnik.

Tam nato digitalizate transkribiramo. Pri tem pa nam program nudi implementirana orodja, ki so nam v pomoč pri tem.

Primer take pomoči je ukaz „Layout Analysis“ oziroma analiza postavitve, kjer program iz digitalizata (slika, .jpg, .tiff…) sam določi kje se nahaja besedilo in ga ustrezno označi.

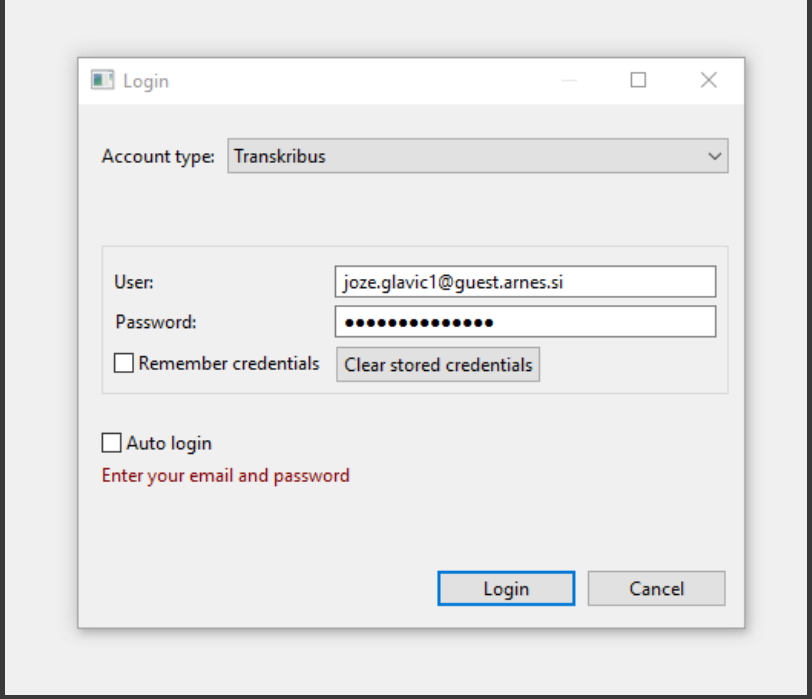
Nadaljnja pomoč, ki jo program omogoča je tudi t.i. Handwritten Text Recognition (HTR) oziroma prepoznavanje ročno zapisanega teksta. Pri tem lahko z ustreznimi predpripravami izvedemo tudi avtomatsko transkribiranje.

UPORABA ZA TRANSKRIBIRANJE

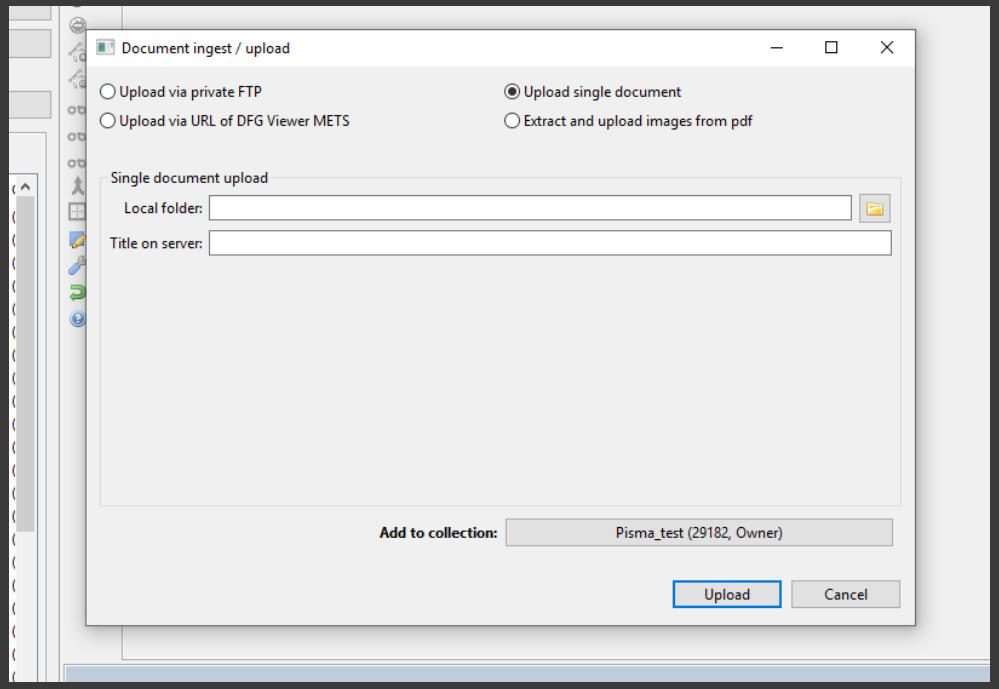


Primer začetnega zaslona programa.

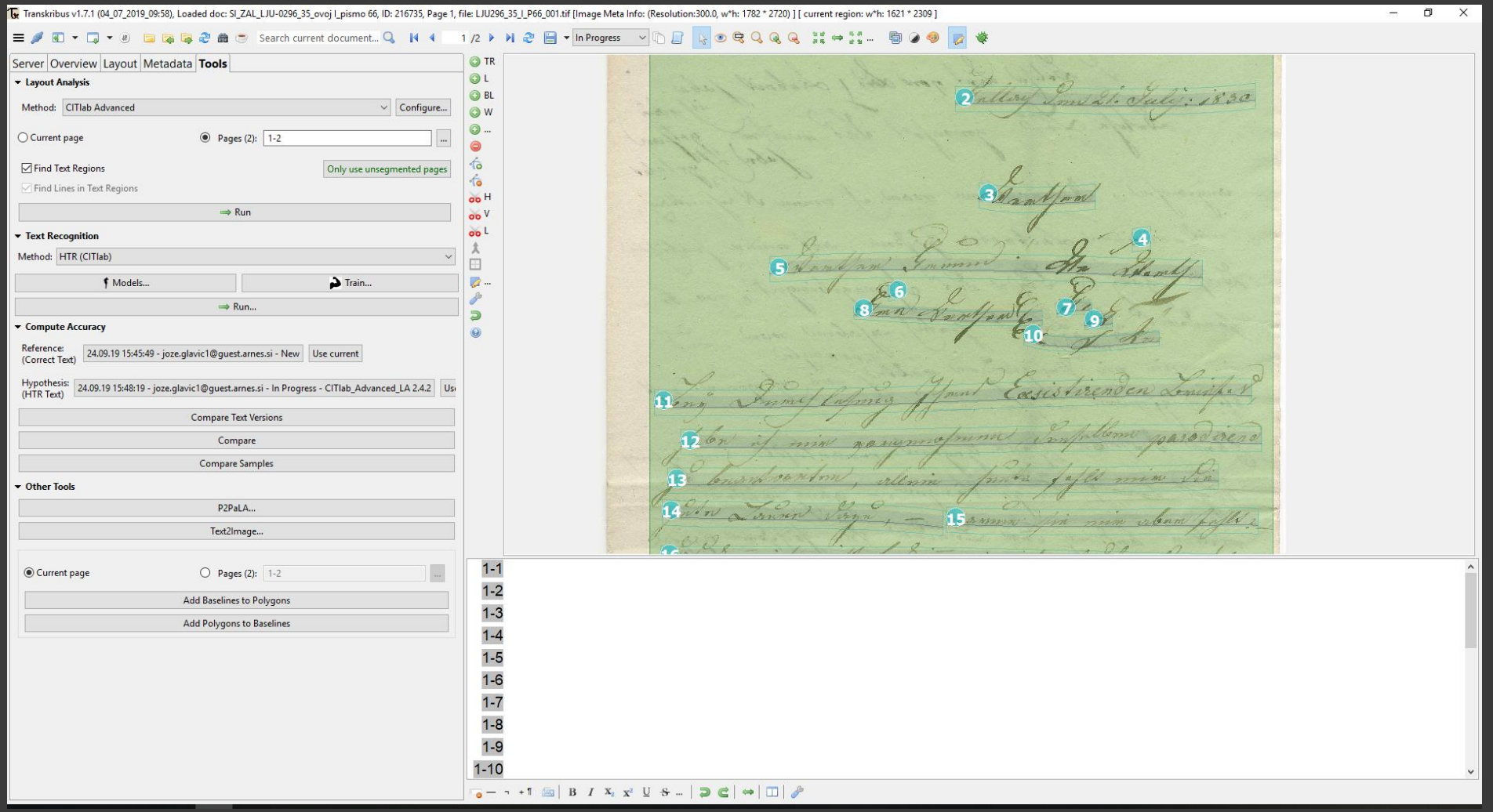
POSTOPEK



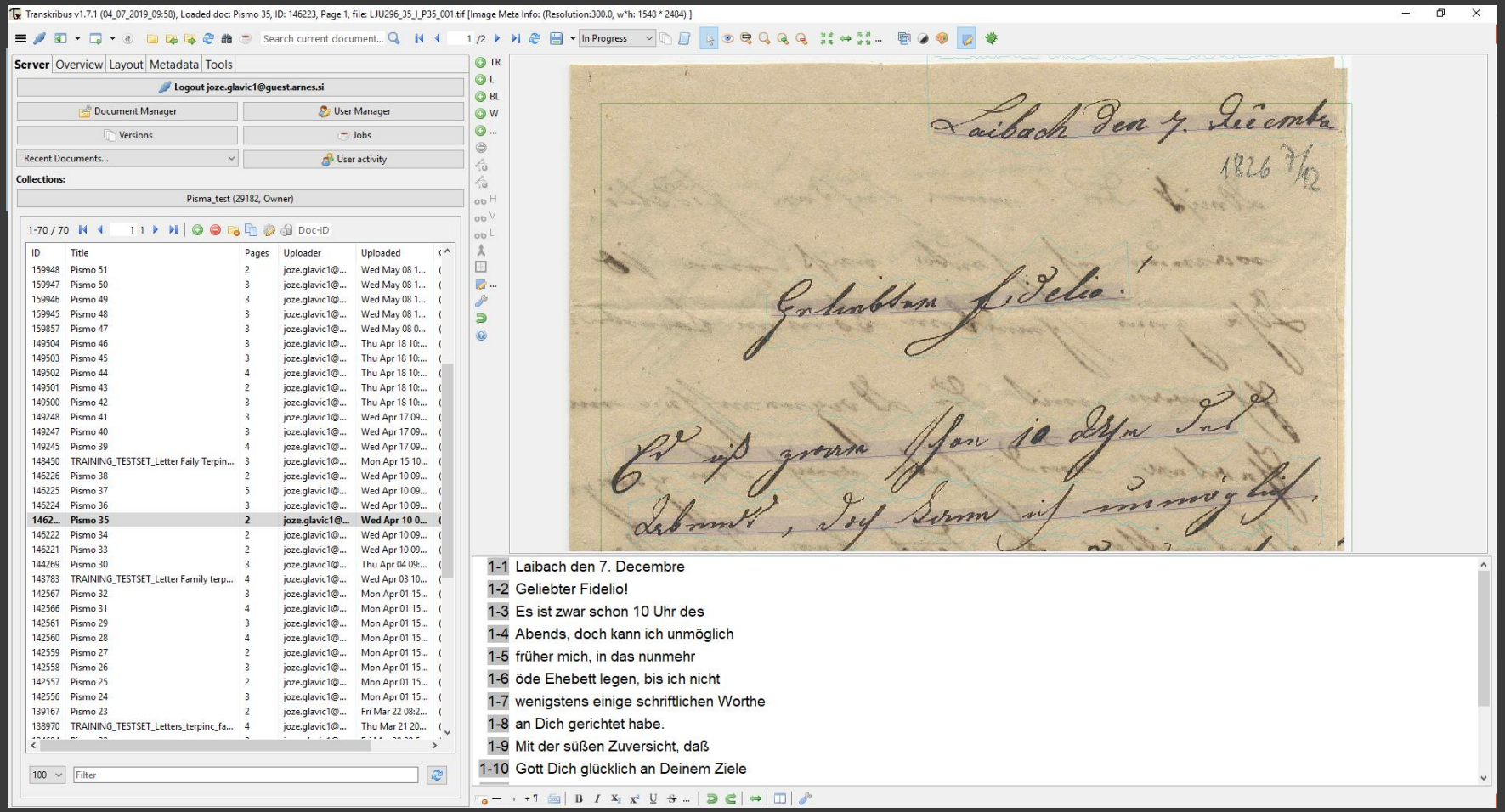
1. Pred prvo uporabo se moramo registrirati.



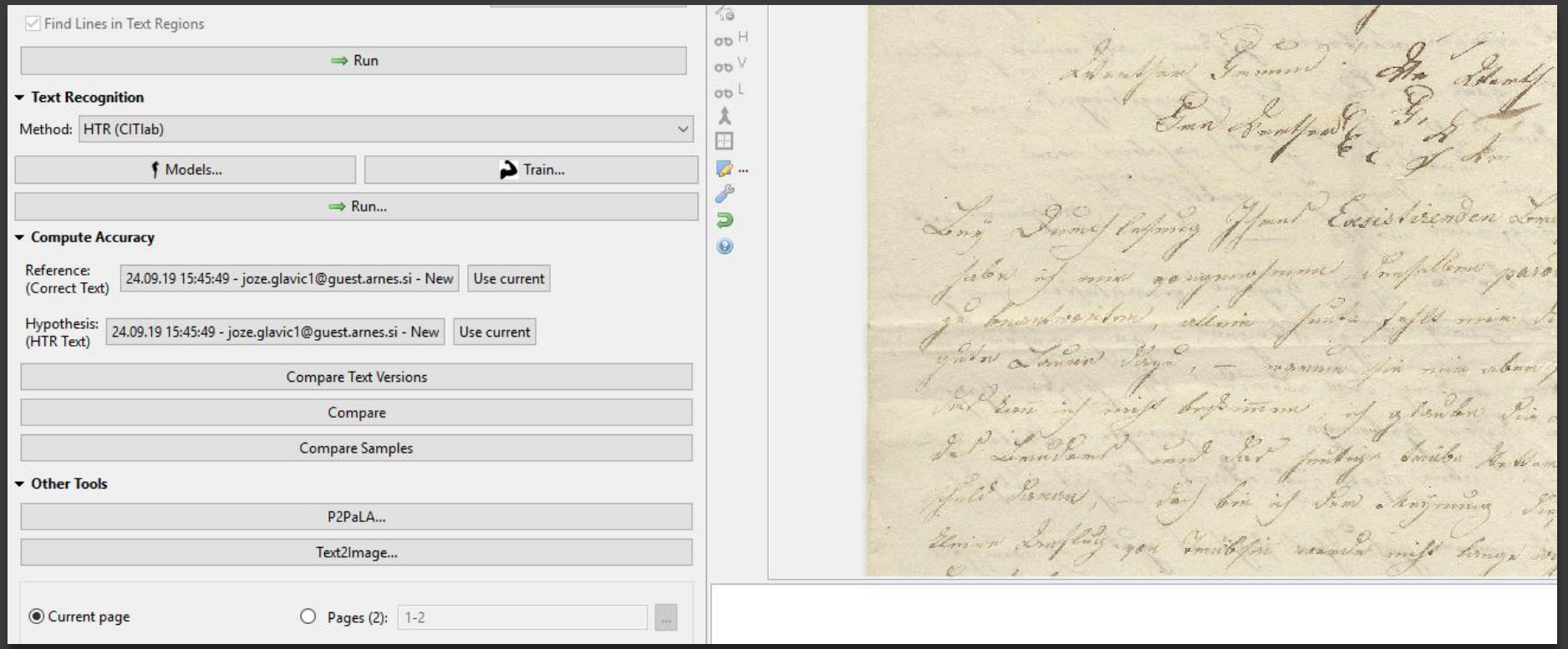
2. Uvozimo digitalizate arhivskega gradiva, katere želimo tranksribirati



Prikaz nastavitev, kjer lahko zaženenmo t.i. ukaz „Layout Analysis“ oziroma analiza postavitve, kjer program iz digitalizata (slika, .jpg, .tiff…) sam določi kje se nahaja besedilo in ga ustrezno označi.



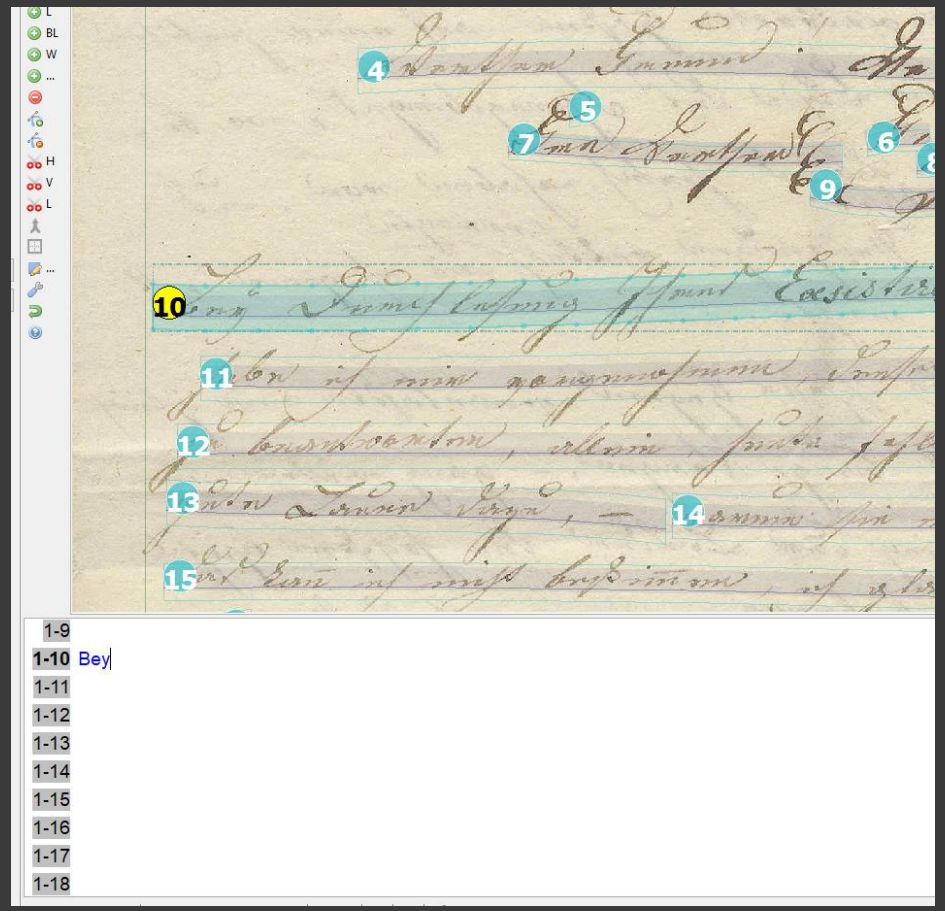
Primer uvoženega digitalizata in njegova obdelava. Na levi strani vidimo seznam uvoženih digitalizatov, na zgornji levi strani sam digitalizat na katerem je program prepoznal besedilo in ga označil. Na spodnji levi strani pa je prikazano ročno transkribiranje digitalizata.



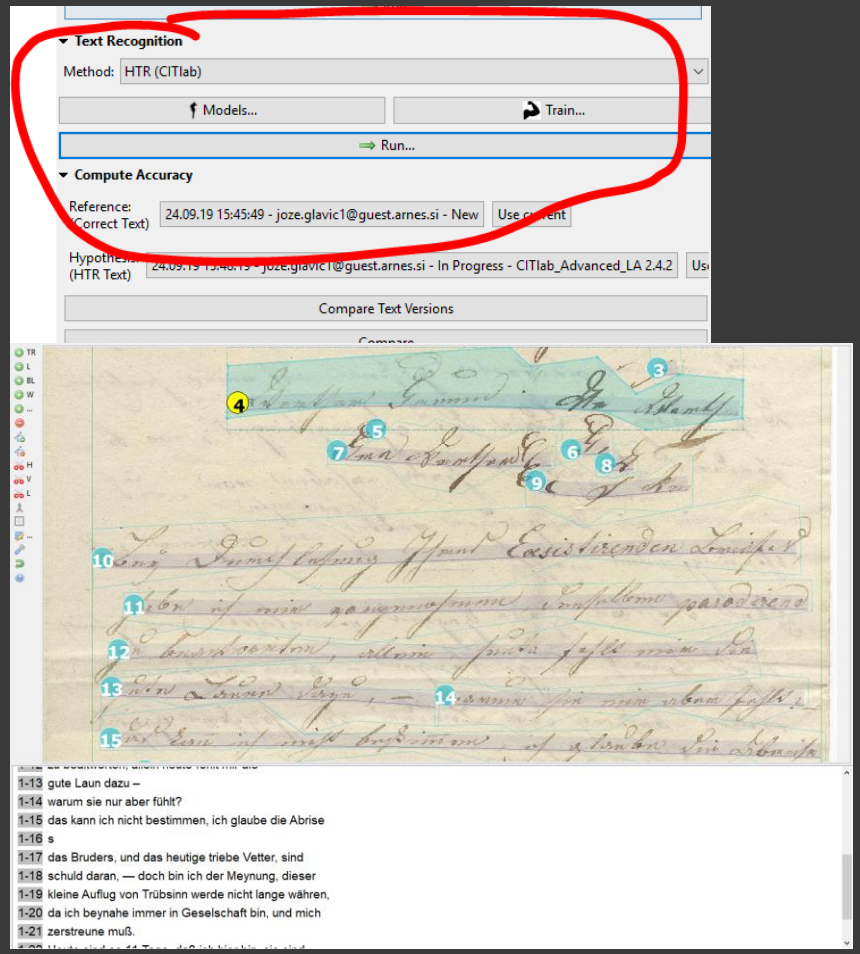
Prikaz možnosti obdelave digitalizata. Tu vidimo, da omogoča program tudi izgradnjo t. i. HTR modela, ki omogoča nadaljnjo avtomatsko transkribiranje digitlaizatov

Gradivo je lahko transkribirano na dva načina. Ali ročno z vpisovanjem uporabnika Transkribusa v sam program. Ali avtomatsko (pri tem je potrebno predhodno izgraditi +HTR model)

Ali ročno



Ali avtomatsko



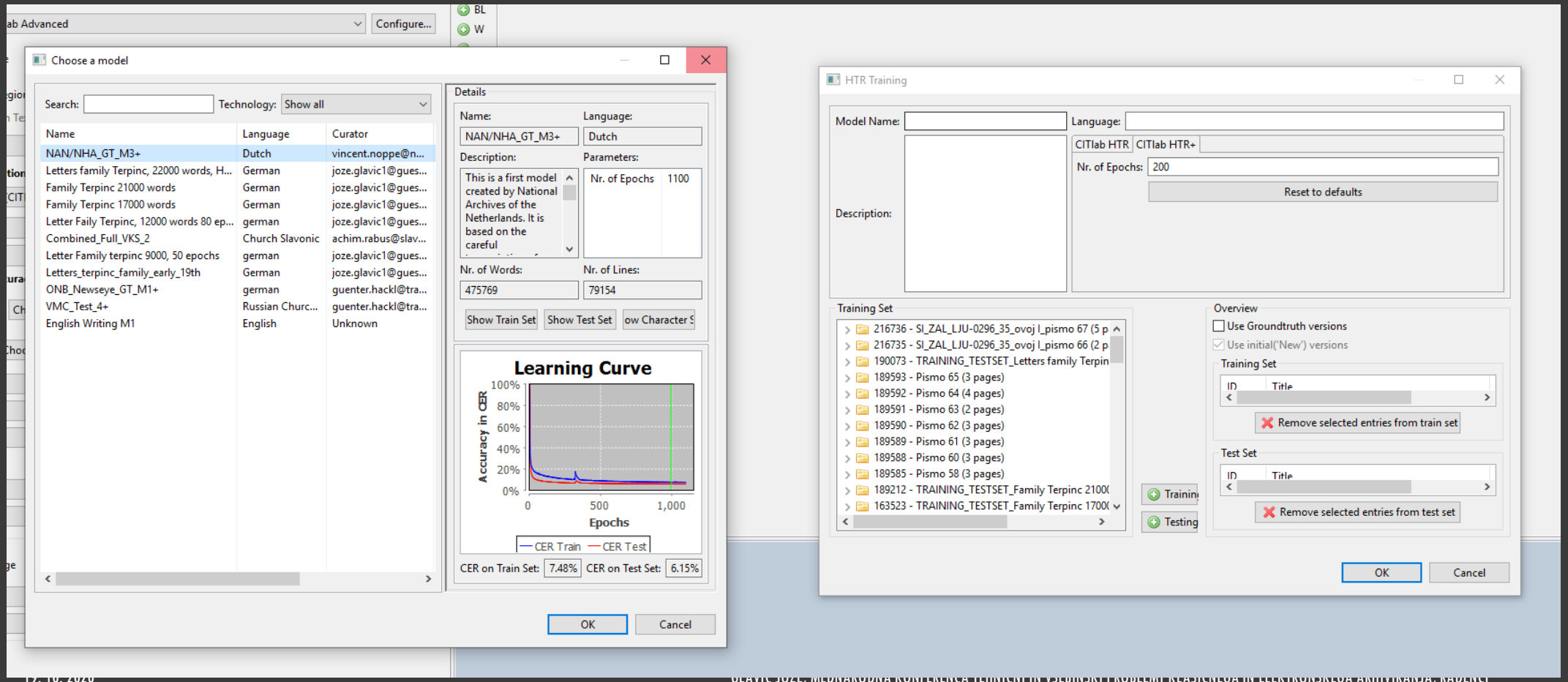
IZGRADNJA MODELA - NA PRIMERU KORESPONDENCE DRUŽINE TERPINC

V Zgodovinskem Arhivu Ljubljana hranimo korespondenco družine Terpinc. Pisma so del večjega fonda Graščina Fužine (SI_ZAL_LJU/0296). Sama korespondenca vsebuje pisma med sorodniki in prijatelji ter tudi nekaj poslovnih pisem. Pisma in korespondente je v 41. izdaji Virov podrobneje opisala in umestila N. B. Kodrič (2018), jih tudi uredila, transkribirala vire ter napisala spremno študijo in kazalo.

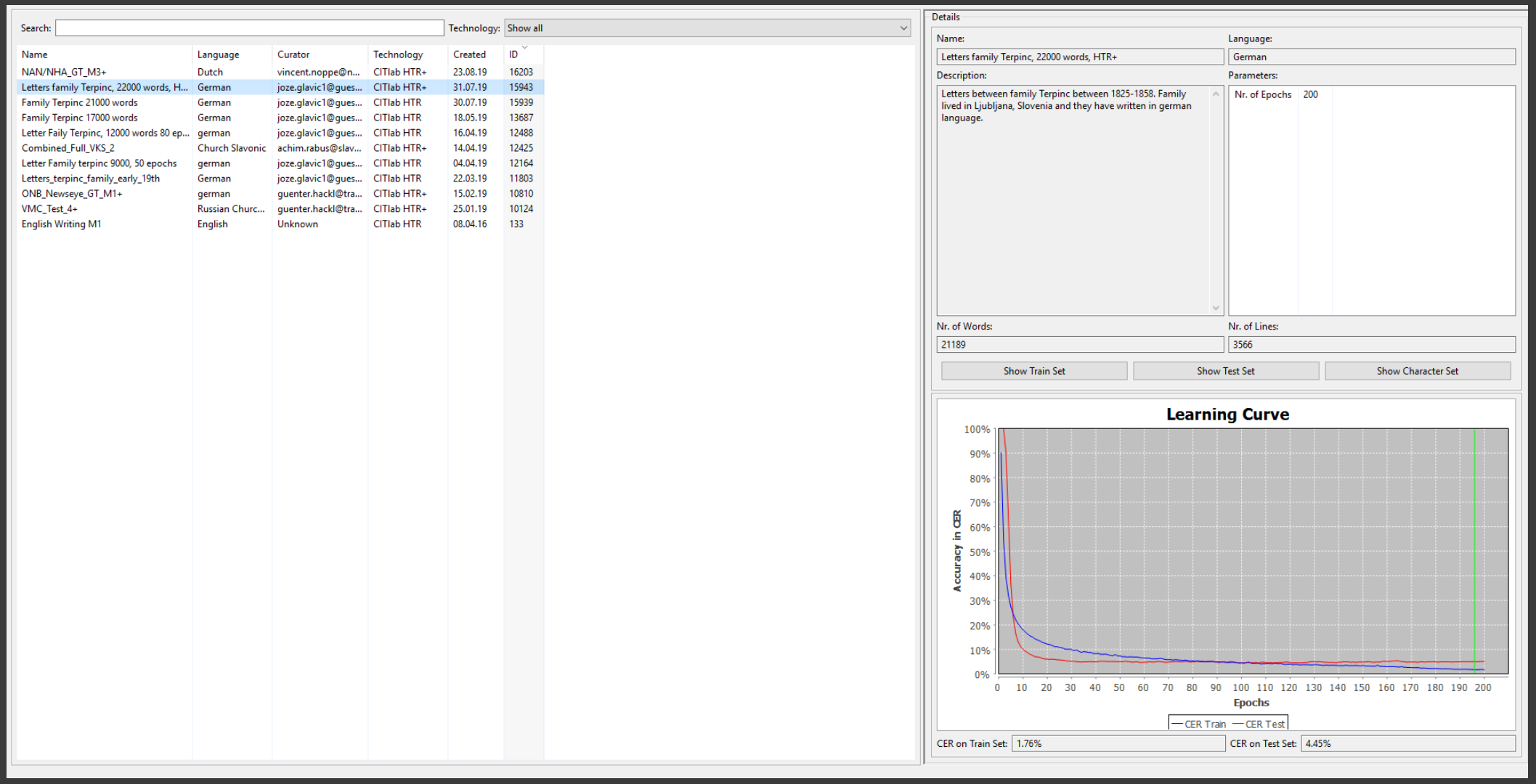
Zaradi velike količine pisem (preko 300) potrebnih transkribiranja, je bilo smiselno preizkusiti izgradnjo modela HTR, ki omogoča delno avtomatsko transkribiranje pisem.

KAJ JE T.I. MODEL ?

Zbirka dokumentov in prepisov(transkripcij) na podlagi katerih se je izvedlo usposabljanje računalniškega programa za optično prepoznavanje



Za izgradnjo modela na korespondenci pisem, je bilo uporabljenih 22000 besed. Program je šel čez 200 epoh (ponovitev). Sam model omogoča avtomatsko transkribiranje pisem z 95.55% natančnostjo. Pogoj seveda je da se uporabijo pisma, katere so pisma istovrstnih avtorjev.



IZGRADNJA MODELA - POGOJI

Za usposabljanje HTR modela je potrebno imeti zbirko digitalizirano dokumentov, ki vključuje od 5000-15000 besed.

Smiselno je, da je gradivo večje količine, da se po izgradnji modela le-ta uporablja.

Potrebno je uporabiti dokumente, ki vsebujejo istovrstno pisavo (so napisani s strani istega avtorja ali vsebujejo slogovno podobno pisavo).

TRANSKRIBUS KOT ORODJE ZA SAMOSTOJNO TRANKSRIBIRANJE

Transkribus se lahko uporabi tudi kot samo pomagalo za ročno transkribiranje. Brez izgradnje modela

Zakaj

- Priročen program, ki olajša delo (dvojni zaslon)
- Standardizirano transkribiranje
- Dodatni znaki so že vključeni (ü-پ﵀خ̣ꞪꞬ₮¦§)

IZVOZ IN KAM NAPREJ

Omogoča izvoz v različne formate (.xml,. pdf, .txt) in podporo standardom (TEI, METS)

Ti formati in standardi, predvsem format .xml podprt s standardom TEI so lahko dobri osnovni gradniki za naslednji korak -> objava nas spletu

Z izvozom v .xml format lahko enostavneje kreiramo digitalne izdaje. Te so lep primer rešitve, kako narediti gradivo bolj dostopno uporabnikom.

Primeri digitalnih edicij:

- https://egov.stmk.gv.at/stmk.gv.at/lbstdig/kategorie/nachlaesse/rosegger-peter/b-korrespondenz
- https://www.gda.bayern.de/DigitaleSchriftkunde/1000_bayhsta_KI-Tegernsee_7.html#_a__b__c__d__e_

Hvala za pozornost!