

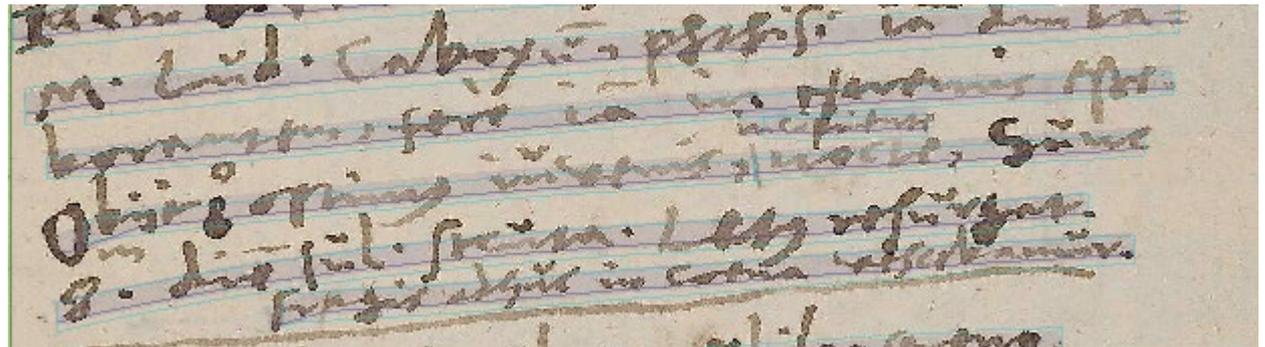


OCR-BW

Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen



Texterkennung von Handschriften mit Transkribus



09.06.2021, Dorothee Huff



Transkribus

- Plattform zur Transkription handschriftlicher und gedruckter Dokumente, die verschiedene Tools vereint, wie z. B. Layouterkennung, Texterkennung per HTR auf Grundlage von neuronalen Netzen, Keyword Spotting
- 2016-2019 Teil des EU geförderten Projekts READ (Recognition and Enrichment of Archival Documents)
- ab Juli 2019 Überführung in READ-COOP SCE
- UB Tübingen seit Januar 2020 Mitglied (aktuell 86 Mitgl. weltweit)
- über 50 000 registrierte Nutzer
- seit Okt. 2020 Erhebung von Nutzungsgebühren



Bearbeitete Quellenkorpora

- Tagebücher der Tendaguru-Expedition des Tübinger Geologen und Paläontologen Edwin Hennig (1897-1973)
- Lateinische Tagebücher des Tübinger Altphilologen Martin Crusius (1573-1604)
- Griechische Predignachschriften von Martin Crusius (1563-1604)
- ausgewählte Bände der juristische Konsilien (1602-1879)
- ausgewählte Bände der Senatsprotokolle (1524-1912)
- Handschriften und Drucke in Malayalam
- Inkunabeln

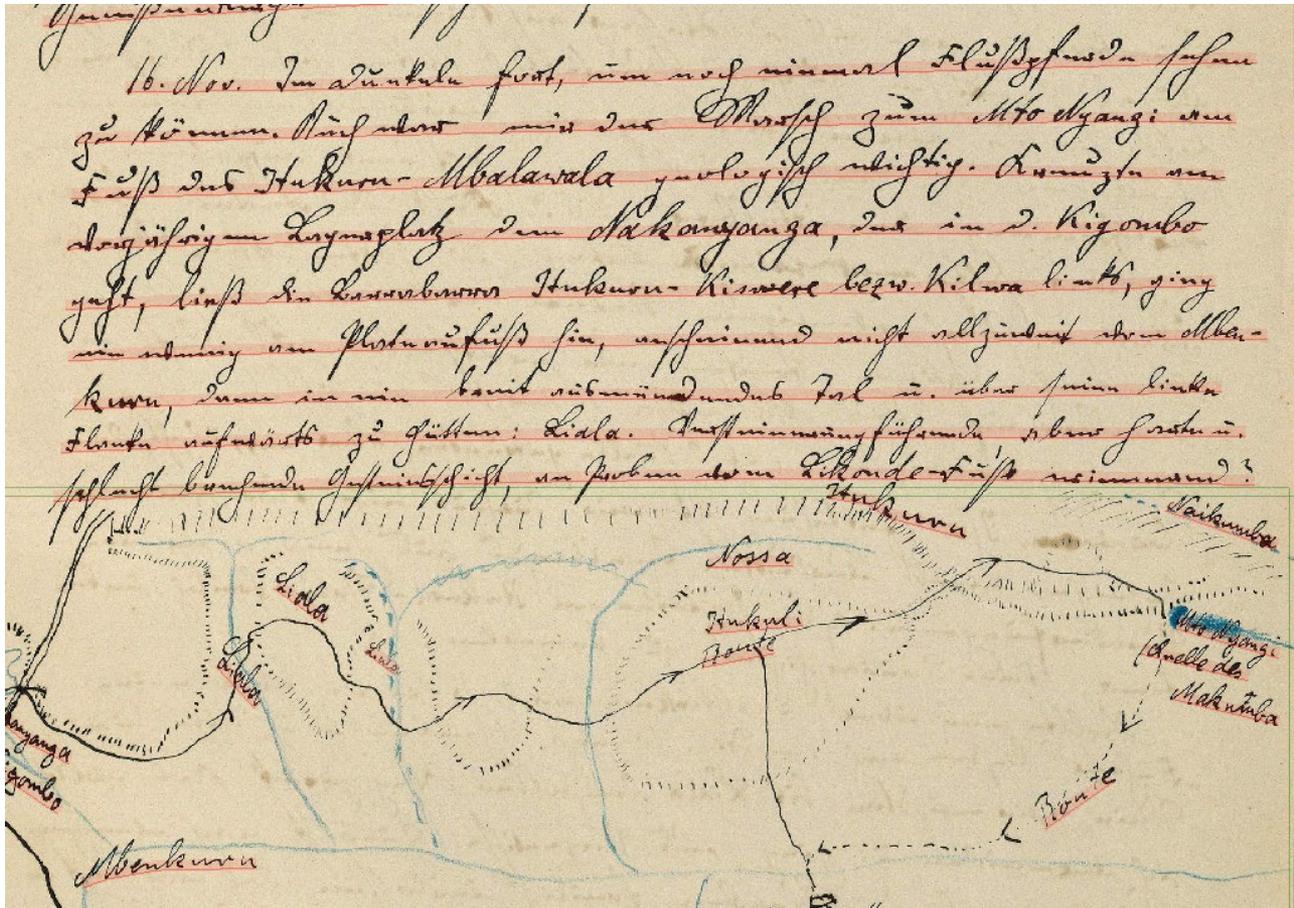


Aufgabe

- Evaluierung von Transkribus anhand verschiedener Textkorpora aus dem Bestand der UB Tübingen
 - Erstellung von Ground-Truth-Daten
 - Modelltraining (Ziel: CER < 5%)
- Vorannahme: für ein gutes Ergebnis muss das Material einheitlich sein



Tagebuch Edwin Hennig, UAT 407/81 (1910-1911)





Tagebücher von Edwin Hennig (1897-1973)

Herangehensweise: Erstellung von GT bis eine CER von unter 5% erreicht ist

Vorläufiges Ergebnis:

- M1: 100 Seiten Ground Truth (1910-1911) → 8.88% CER
 - M2: 133 S. GT (1910-1911) → 5,58% CER
 - M3: 165 S. GT (1910-1911) → 4,32% CER
-
- **Problem:** das Modell performte auf früheren oder späteren Jahrgängen der Tagebücher deutlich schlechter



Modellentwicklung

	UAT_M1	UAT_M2	UAT_M3	UAT_M4	UAT_M5	UAT_M6	UAT_M6/La anguageMod el		Comb_Deut sch_Latein_ XX_M2	German_Ku rrent_XIX_ M1
S. 1 (1910-11)	2,97	2,03	1,88	1,88	2,03	1,74	1,38		18,62	14,13
S. 2 (1910-11)	3,09	3,09	2,7	2,08	2,47	1,85	1,62		15,9	12,73
S. 3 (1910-11)	18,7	17,58	17,36	16,69	16,01	15,23	14,39		30,85	26,6
S. 4 (1910-11)	8,96	9,3	9,3	8,69	8,96	8,69	8,08		21,11	18,81
S. 5 (1910-11)	1,92	1,37	0,89	1,24	0,82	1,1	0,89		18,56	12,65
S. 6 (1910-11)	2,2	1,52	1,24	1,24	1,52	1,31	1,17		18,8	13,09
S. 7 (1910-11)	3,15	1,8	1,75	1,97	1,69	1,63	1,52		15,26	11,94
S. 8 (1910-11)	2,12	1,33	1,27	1,03	1,39	1,15	0,97		15,73	12,22
S. 9 (1911)	2,07	1,31	0,98	0,92	1,03	0,98	0,92		14,25	11,21
S. 10 (1944-46)	18,78	19,5	17,01	8,09	6,74	6,64	5,71		31,22	29,46
S. 11 (1944-46)	11,72	8,54	8,39	3,4	3,04	3,4	2,97		23,01	20,26
S. 12 (1961-62)	16,72	14,37	12,35	8,05	2,64	2,64	2,64		19,36	16,79
S. 13 (1961-62)	20,8	16,79	15,68	11,42	3,52	2,9	2,72		22,65	17,84
S. 16 (1897)	57,69	53,95	49,64	46,19	35,7	6,54	4,53		9,27	8,05
S. 15 (1899)	30,64	29,95	23,7	22,64	16,23	4,27	3,35		14,71	9,6
S. 16 (1901)	14,17	12,31	11,66	10,45	8,26	4,62	4,45		18,54	13,2
Durchschnitt	8,88 (12,94)	5,58 (11,6)	4,32 (10,52)	4,31 (8,86)	4,03 (6,81)	4,05	3,61		23,3	15,36
Seiten im Traini	95	126	156	184	212	224				
Testseiten	5	7	9	11	13	16				
GT insgesamt	100	133	165	195	225	240				



Lösung: Nachtraining mit Material aus weiteren Jahrgängen

- M4: 165 (1910-1911) + 30 (1944-1946) → 4,31% CER
- M5: 165 (1910-11) + 30 (1944-46) + 30 (1961-62) → 4,03% CER
- M6: 165 (1910-11) + 30 (1944-46) + 30 (1961-62) + 15 (1897-1901) → 4,05% CER (mit Language Model 3,61% CER)

Erkenntnis: auch zeitlich heterogenes Material kann mit genauso gutem Ergebnis in einem Modell vereinigt werden

Annahme: dies sollte auch für Dokumente mit unterschiedlichen Schreibhänden gelten

→ entsprechend Anpassung der Projektplanung



M1: CER 2.07%, WER 8.53%

8

Büffel. Sandte Leute nach ihnen aus u. erhielt um 4 etwa Meldung. Darauf ging ich mit starker Begleitung bis zu ~~den-dem~~ überraschend entfernten Fluß (ca ½ Std. Die Leute schöpfen z. T. aus Wasserlöchern!), u. jenseits ~~im-in~~ d. Dickicht. ~~Plötz., Plötz.,~~ lich ging unmittelbar vor uns d. Gekrache los, wir ~~waren-ahn waren, ehe ich nach-noch~~ d. Gewehr selbst genommen hatte, der ~~starkein-starken~~ Herde auf vielleicht 25 m. ~~nate-nahe~~ gekommen, ohne daß auch nur das ~~Geriegste-Geringste~~ v. d. Tieren sichtbar wurde. Wir flüchteten ~~eilichst-eiligst~~, da sie ebensogut auf uns ~~zukommen-zukommen~~ konnten. Auch Brüllen der Kühe ertönte. Da hier nicht möglich war, zu Schuß zu kommen, wollte ich in eine angrenzende ~~gangwa-saffenen-yangwa~~ (offenen Wald) auf die sie zu geflüchtet waren. Meine Führer gingen furchtlos, wenn auch vorsichtig in gleicher Richtung ~~weiter-weiter~~, wobei plötzlich auch auf d. andern Seite d. Krachen losging. Das war mir doch etwas unheimlich, zumal jede Kugel nach spätestens 5 m. abge-, lenkt werden mußte u. so nahe wohl nur ein bereits an-, ~~stärrender-stürmender~~ Büffel kommt. Wir gingen dann am Rande rings um d. Gehölz, ohne etwas zu ~~sehen-sehen~~ (als die Wald-bestandene Fläche, die hier ~~anscheinend-anscheinend~~ beim Zusammenstoß der Täler d. ~~Maruai-Mavudyi~~ u. ~~Kikende-Kihendye~~ in den Busch u. d. Hügelland unvermittelt eingestreut ist.) Auf d. Weg zurückgekehrt trafen wir den ~~Jumben-Jumben~~, einen alten freundlichen hübschen ~~Grnis-Greis~~ mit 2 kleinen ~~Kindern-Kindere~~ ruhig auf d. Nachhausewege, er berichtete daß die Spuren nahe am Fluß ü. d. Weg führten. In d. Tat ~~wart~~ war die ganze Herde im Bogen ~~rückwärts durchgebrochen-rückwärts durchgebrochen~~ u. zwar im übrigen völlig lautlos. Es war überflüssig in dem Dickicht zu folgen, ich kehrte wieder unverrichteter Sache ~~heim-heim~~, aber doch um ~~einige-seltsam-einige-seltsame~~ Erfahrungen reicher. Die ~~Leuthen-Leuthen~~ hier in diesem verlorenen Talkessel kamen mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.

4. Aug. Rückkehr nach ~~Makangaga-Makangaga~~. Morgens schöne Wanderung

M6: CER 0.92%, WER 3,47%

8

Büffel. Sandte Leute nach ihnen aus u. erhielt um 4 etwa Meldung. Darauf ging ich mit starker Begleitung bis zu dem überraschend entfernten Fluß (ca ½ Std. Die Leute schöpfen z. T. aus Wasserlöchern!), u. jenseits ~~im-in~~ d. Dickicht. Plötz., lich ging unmittelbar vor uns d. Gekrache los, wir ~~waren-waren~~, ehe ich noch d. Gewehr selbst genommen hatte, der starken Herde auf vielleicht 25 m. nahe gekommen, ohne daß auch nur das Geringste v. d. Tieren sichtbar wurde. Wir flüchteten ~~nlichst-eiligst~~, da sie ebensogut auf uns zukommen konnten. Auch ~~Brüllen-Brüllen~~ der Kühe ertönte. Da hier nicht möglich war, zu Schuß zu kommen, wollte ich in eine angrenzende yangwa (offenen Wald) auf die sie zu geflüchtet waren. Meine Führer gingen furchtlos, wenn auch vorsichtig in gleicher Richtung ~~weiter-weiter~~, wobei plötzlich auch auf d. andern Seite d. Krachen losging. Das war mir doch etwas unheimlich, zumal jede Kugel nach spätestens 5 m. abge-, lenkt werden mußte u. so nahe wohl nur ein bereits an-, stürmender Büffel kommt. Wir gingen dann am Rande rings um d. Gehölz, ohne etwas zu sehen (als die Wald-bestandene Fläche, die hier ~~anscheinend-anscheinend~~ beim Zusammenstoß der Täler d. Mavudyi u. ~~Kihendze-Kihendye~~ in den Busch u. d. Hügelland unvermittelt eingestreut ist.) Auf d. Weg zurückgekehrt trafen wir den ~~Jumben-Jumben~~, einen alten freundlichen hübschen ~~Grnis-Greis~~ mit 2 kleinen ~~Kindern-Kindere~~ ruhig auf d. Nachhausewege, er berichtete daß die Spuren nahe am Fluß ü. d. Weg führten. In d. Tat war die ganze Herde im Bogen rückwärts durchgebrochen u. zwar im übrigen völlig lautlos. Es war überflüssig in dem Dickicht zu folgen, ich kehrte wieder unverrichteter Sache ~~heim-heim~~, aber doch um einige seltsame Erfahrungen reicher. Die Leuthen hier in diesem verlorenen Talkessel ~~kannen-kamen~~ mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.

4. Aug. Rückkehr nach Makangaga. Morgens schöne Wanderung



x von hier be-
 stiegen wir
 noch den Halm,
 einen Berg,
 von dem mag
 eine sehr schö-
 ne Aussicht
 auf die See sein.

 arbeitete. Um $\frac{1}{2}$ 1 gingen wir nach Kurth's
 Gasthaus und aßen dort sehr gut zu
 Mittag. Von dort gingen wir wieder nach
 Hause und waren unschlüssig, ob wir ge-
 hen sollten, oder nicht. Ich arbeitete noch
 etwas und, da es nur ganz wenig regnete,
 wollten wir zum Förster gehen. Als wir
 ein Stück die Chaussee gegangen waren,
 fing es an stärker zu regnen, doch gingen
 wir weiter und es hörte bald wieder auf.
 Wir kehrten in einem großen Restaurant
 mit einer Kapelle ein, das mir gar nicht
 sehr förstermäßig vor kam, da auch Kellner
 da waren. Nachher gingen wir weiter und
 merkten auch, dass wir gar nicht beim För-
 ster gewesen waren, sondern kamen erst
 jetzt daran vorbei und stiegen auf einen
 Berg mit dem einem Turm war. Dieser war



M1: CER 57.69%, WER 93.73%

(17.)

belete.aerimachdr-beitete. Um ½1 gingen wir nach Kurth's
Barlharvne afindrrsvhrhgabas-Gasthaus * und aßen dort sehr gut zu
Nitbag. Ba dors gmmrarmibeerrach. Mittag. Von dort gingen wir wieder nach
Nlaave nad raren ihlarhbwir Gle-Hause und waren unschlüßig, ob wir ge-
kenollbenoerast. Sharbaibobenreh-hen sollten, oder nicht. Ich arbeitete noch
Froaecears vr Mrr wer rembl- etwas und, da es nur ganz wenig regnete,
KolbbienrurwBndlrgeser. Lbwir- wollten wir zum Förster gehen. Als wir
aa-Mik- ein Stück die Shaaaa-Geerepraarer- Chaussée gegangen waren,
EineDa-Bleeskerarlaen- dochaga- fing es an stärker zu regnen, doch gingen
Moraeebonne- eibiolebalear- wir weiter und es hörte bald wieder auf.
Pirkehrlern- Maemroen-Nalauruk-. Wir kehrten in einem großen Restaurant
inenPapellee- Nagrrarer- mit einer Kapelle ein, das mir gar nicht
kbermjirirbara, Laaachiellari- sehr förstermäßig vorkam, da auch Kellner
NawaremNalkaoroaaamaraerna- dawaren. Nachher gingen wir weiter und
kerkiendaei- daibeiaarrabbeiForu- merkten auch, daß wir garnicht beim För-
ler. Gererer worvo, Wadera- Kareners- ster gewesen waren, sondern kamen erst
d- MiabDaraa- orbemababiegeauberer- jetzt daran vorbei und stiegen auf einen
Berye- mrrdura- Berg, auf dem einen Turm war. Bieseraer- Dieser war
Aaber- rerseklosseruadrar- Neeenaden- aber verschlossen und wir stiegen an den
Dbrond- hiab. Dordsmaarein- omekya- Strand hinab. Dort war ein ziemlich gro-
ser- Ber, aber Brnjer- Pandberabesaus- junger Hund, der alles aus dem
Maverkeraushalae- rraswankrar- Wasser herausholte, was man hineinwarf,
brih- Naerrihmrrk- korile- doch da er nicht schwimmen konnte,
Donderrrdowers, aberBaraahake- ging er nur so weit, als er Grund hatte,
Nac-, abenkeererbruaererber- Waro- und, als ein kleinerer Hund weiter hinaus-
awarra- Nararo- Nedisch- Nacer- Aer- schwamm, war er so neidisch, daß er den
Mdr- rbweairberawrrarererlebe- andern nicht wieder herauslassen wollte
balabaraikbarrerjaelrba- wellen- und ihn nachher verjagte. Als wir weiter
denilaiberakelBorne- ngenaben- der Kaiser Wilhelm Brücke zugingen, sahen
esenraberollen- Berenbogerra- wir einen wundervollen Regenbogen von
Neodrog- a- Neerraensdernde- Misdroy- aus ins Meer ragen, denn die
Grmbwar- Mederkemrracbooojerwaa- Sonne war wieder hervorgekommen und
Kooneker- ben- * vorher be-
zliaBessurir- stiegen wir
roek- Neaala- noch den Kulm,
iela- Beref- einen Berg,
on- Neraz- von dem man

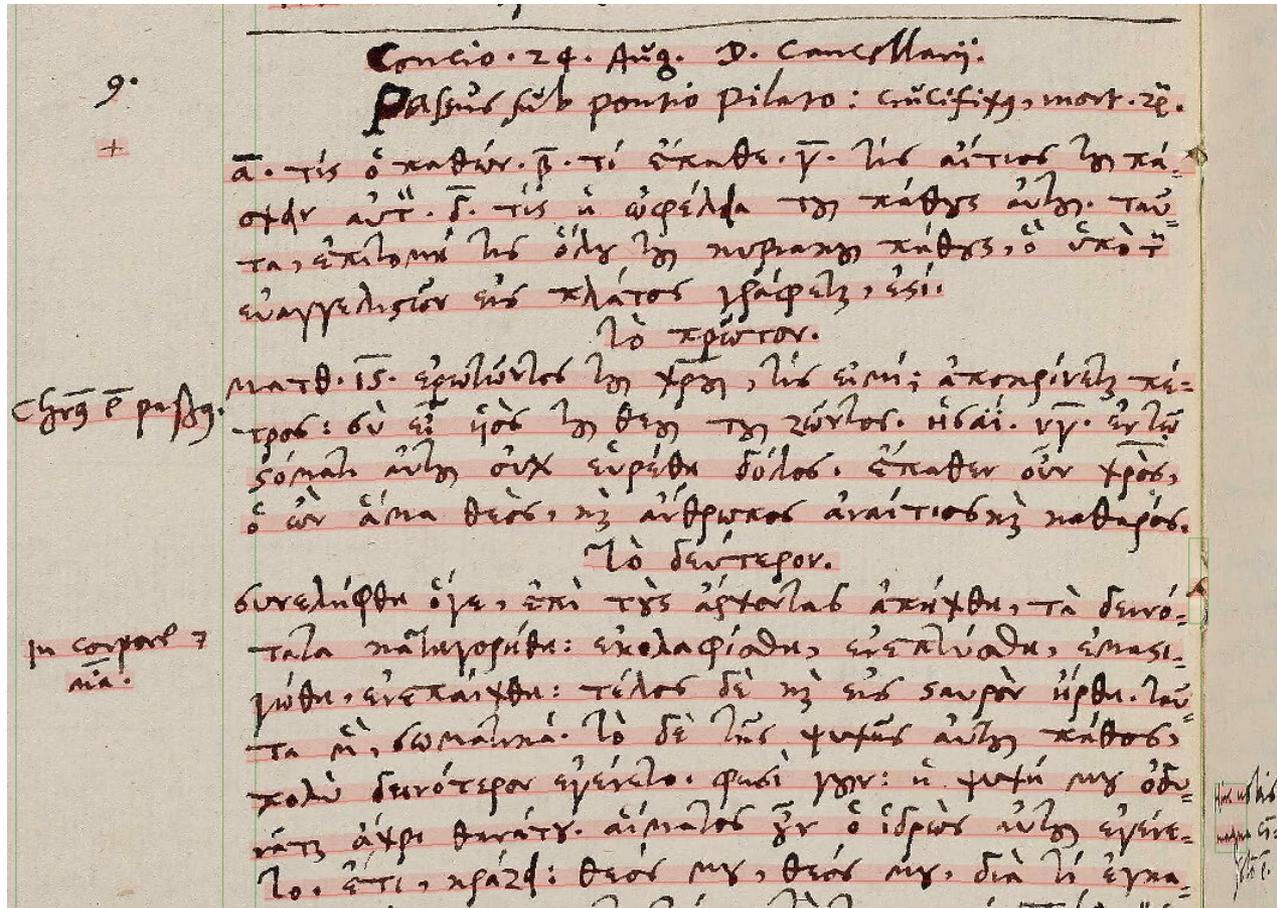
M6: CER 4.53%, WER 18.82%

(9-7.)

beitete. Um ½1 ~~ginden-gingen~~ wir nach ~~Kurtes-Kurth's~~
Gasthaus * und aßen dort sehr gut zu
Mittag. Von dort gingen wir wieder nach
Hause und waren unschlüßig, ob wir ge-
hen sollten, oder nicht. Ich arbeitete noch
etwas und, da es nur ~~gant-ganz~~ wenig regnete,
wollten wir zum Förster ~~gesen-gehen~~. Als wir
ein Stück die ~~Chaussee-Chaussée~~ gegangen waren,
fing es an stärker zu ~~rechnen-regnen~~, doch gingen
wir ~~weiten-weiter~~ und es hörte bald wieder ~~aus-auf~~.
Wir kehrten in einem großen ~~Restauran-Restaurant~~
mit einer Kapelle ~~ein-ein~~, das mir gar nicht
sehr ~~förstermäßig vor-kam~~, ~~förstermäßig~~ vorkam, da auch ~~Kellzer-Kellner~~
dawaren. Nachher gingen wir weiter und
merkten auch, daß wir garnicht beim ~~Hor-För-~~
ster gewesen waren, sondern kamen erst
jetzt daran vorbei ~~undstiegen-und~~ stiegen auf einen
Berg, auf dem einen Turm war. ~~Sieser-Dieser~~ war
aber verschlossen und wir stiegen an den
Strand hinab. Dort war ein ziemlich gro-
~~ßer-Ber~~, aber junger Hund, der alles aus dem
Wasser herausholte, was man ~~hinein-warf~~,
doch da er nicht schwimmen ~~konnte-konnte~~,
~~gind-ging~~ er nur so weit, ~~es-als~~ er Grund hatte,
und, als ein kleinerer Hund weiter ~~hinaus-hinaus-~~
schwamm, war er so neidisch, daß er den
ändern nicht wieder ~~herausassen-herauslassen~~ wollte
und ihn nachher ~~vergagterts-verjagte~~. Als wir ~~weiten-weiter~~
der ~~Haiser-Kaiser~~ Wilhelm Brücke ~~zu-gingen~~, zugingen, sahen
wir einen wundervollen ~~Gegenbogen-Regenbogen~~ von
~~Misdrog-Misdroy~~ aus ins Meer ragen, denn die
~~Sound-Sonne~~ war wieder ~~henvorgekommen-hervorgekommen~~ und
~~Kvenher-bei-~~ * vorher be-
stiegen wir
noch den ~~Duhm-Kulm~~,
einen ~~Bergt-Berg~~,
von dem ~~Tag-man~~



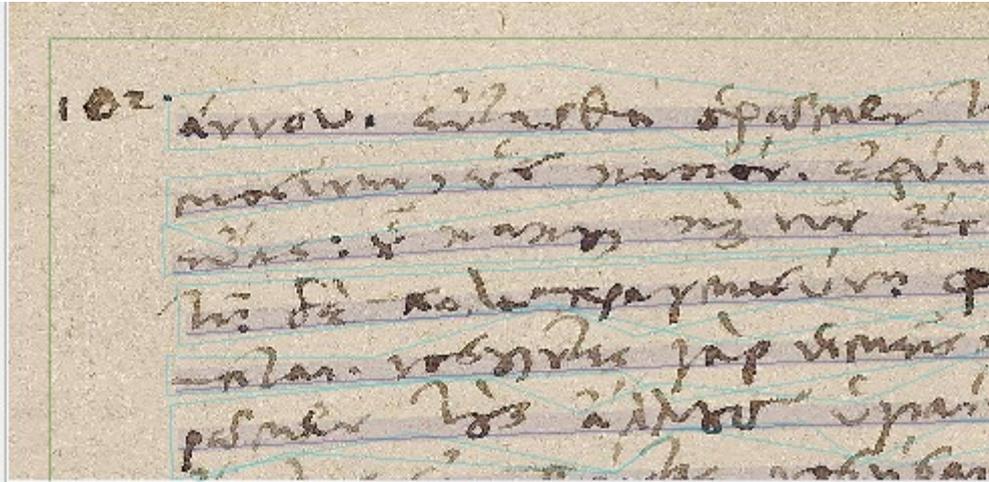
Griech. Predignachschrift, Mb 19-4, (1563-70)





Griechische Predignachschriften von Martin Crusius

- **Herangehensweise:** jeweils 10 Seiten GT aus 20 Handschriften für Modelltraining (18 Bände Predignachschriften + 1 Martyrologium von Crusius, 1 Band Predignachschriften von Samuel Grammer)
- **Ergebnis:** CER von durchschnittlich 3,41% für die griechischen Textteile der 18 Bände Predignachschriften von Martin Crusius (→ insgesamt auf über 10.000 Seiten anwendbar)
- **Nächster Schritt:** Vereinigung mit dem Modell für die lateinischen Tagebücher von Martin Crusius



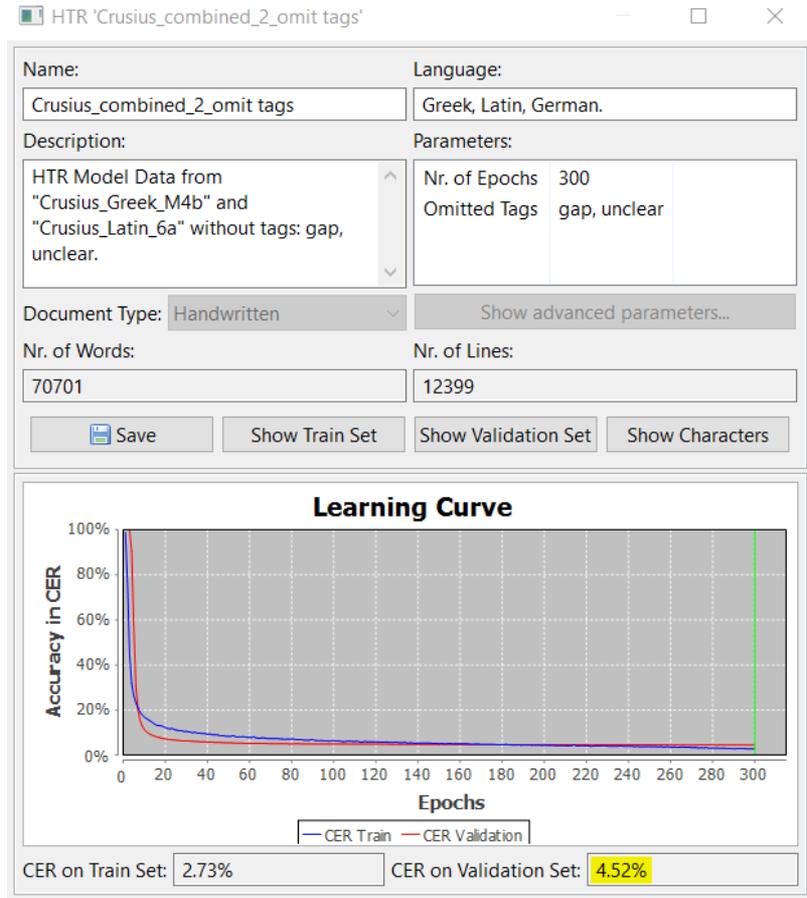
Mb 19-1, S. 58 (1587)	
Mb 19-2, S. 145 (1566)	2,56
Mb 19-3, S. 118 (1564)	3,66
Mb 19-4, S. 107 (1564)	2,56
Mb 19-7, S. 484 (1575)	3,52
Mb 19-8, S. 14 (1575)	3,64
Mb 19-9, S. 188 (1577)	6,4
Mb 19-10, S. 12 (1579)	1,93
Mb 19-11, S. 70 (1581)	3,66
Mb 19-12, S. 24 (1582)	3,87
Mb 19-13, S. 12 (1583)	3,08
Mb 19-14, S. 34 (1587)	1,93
Mb 19-15, S. 192 (1589)	3,23
Mb 19-16, S. 612 (1594)	3,54
Mb 19-17, S. 204 (1595)	3,08
Mb 19-18, S. 116 (1597)	3,53
Mb 19-19, S. 55 (1600)	3,76
Mb 19-20, S. 69 (1602)	3,56
Mb 12,	
Mb 17, S. 14 (1594)	2,44
CER im Durschnitt	3,41
GT Training	162
GT Validation	18
GT insgesamt	180

- 1-1 άννου· ένταϋθα· όρωμεν· τήν· πολυπραγ=↵
- 1-2 μοσύνην· ώς· κακόν· έφύη· αύτη· έκ· τής↵
- 1-3 εϋας·· οϋ· κακοϋ· χ'· νϋν· έτι· άπολαύομεν·↵
- 1-4 τή· δέ· πολυπραγμοσύνη· φθόνος· προσή=↵
- 1-5 πται· νοσοϋντες· γάρ· ήμεϊς· οϋχ'· ήδέως· ό=↵
- 1-6 ρώμεν· τοϋς· άλλους· υγαιίνοντας· άλλα· τί↵
- 1-7 όφελος· εί· πάντες· νοσήσαιεν·· οϋδεις· άν· άλλ=↵



Kombinationsmodell Crusius

- 346 S. GT
- Sprachen: Griechisch, Latein, Deutsch
- Schreiber: hauptsächlich Martin
- Crusius
- Zeitraum: 1563-1604





Juristische Konsilien, UAT 84/14 (1659-1661)

Zit.
 also 3. zum Andern Die Laici sind seculars
 juris decimandi nicht capaces was auch
 Decimarum n. jus est spirituale et decima pp
 divina et sacramentorum administrationem
 solvuntur, et ideo laicis illas possidere non
 licet
 cap. quomodo 17. cap. tua. 25. c. dudum. 31. & d. decim.
 Andr. Vallens. de jur. canon. lib. 3. t. 30. § 6. n. 1.
 Ungebaue in coment. Decret. lib. 3. t. 30. n. 8.
 Ubi a. adversus agentem et decimas petentem
 est juris presumptio et dispositio, ibi nec spolia,
 to decima sunt restituenda.
 Caspar. Klock. tom. i. Cons. 40. num. 13.
 Qui dicitur parvitas in dem allecarben Hough
 de 1546. expresse gemeldet, 33 Sprittige
 Zehnen zur Kaszierung in Hingstall Non. Albert
 hies vngis sind alle was gesondt habe; samon.
 jaw an nos magel laben zu Es dan sind nicht
 dellen zu separiran.
 Bona n. et redditus Ecclesiastici sunt Patrimonium
 Christi et ad profanos usus non transferendi
 Caspar. Klock. tom. i. vol. 2. cons. 35. n. 88.
 et seq.



Malayalam, CiXIV131-4 (1877) und Mal 765-1 (1796-1859)

BELUCHISTAN.
ബലൂചിസ്ഥാൻ.

യവനൻ ഗെദ്രോസി എന്നു വിളിക്കുന്ന ബലൂചിസ്ഥാൻ
ന്റെ പൂർവ്വദക്ഷിണ ഭാഗത്തിൽ ഏകദേശം 127000 നാഴ്
യി വിളങ്ങുന്നു. മഹാഅലക്സന്ദർ (325 ക്രി. മു.) ഇന്ത്യയിൽ
പോവഴി ഈ നാട്ടിലൂടെ സൈന്യസഹിതം കടന്നു വേ
വും ഉണ്ണുവുകൊണ്ടു സൈന്യത്തിന്റെ ങ്കാഹരി ചതു
ദിവസം യാത്ര ചെയ്തു ശേഷം മാത്രം തിരിച്ചു പാർസ്യയിൽ
നാടു എത്രയും ഉണ്ണുവും പാഴായ്ക്കും ആകുന്നു. ചിലേടത്
ങ്ങൾ പൂഴിപ്രദേശങ്ങൾ ഇവയല്ലാതെ മറ്റൊന്നും കാ
ല ഇടങ്ങളിൽ മാത്രം ഉറവുകളും അല്പ പുഷ്പിയും ഉണ്ടു,
സികളായ ബലൂചിസ്ഥാൻടെ സംഖ്യ ഏകദേശം 2000000.
നൂറ്റും ശക്തിയും ഉള്ള ജാതി തന്നെ. അറബികളുടെ സ
വർ ബഹു താല്പയ്ക്കത്തോടെ മുഹമ്മദ് മതത്തിൽ ചെന്നു
പോലെ സുനിദർ ആകയാൽ ശീതരായ പാർസികളെ ഏ
ന്നും. അവരുടെ ഭാഷക്കു പാർസ്യഭാഷയോടു അടുപ്പുണ്ടു.

മൽവുലുളു മക...

കെകുടേ മ വെ പ്പ റ ന്ദ ന്ദ ഇൻ വിവസ്
പു റ്റകു കുമം ഉഷ്റ കുമ ഹി ന്ദ മൻ
കു റ കും ഹി റ യ ഉഷ്റ കെ റു കി ക മ
സം ഹൃ ത സ മ് ഉ ന്ദ ഹൃ ക്കു കെ വെ
ജ മ റ സം ഇ ത്ത കെ മ മ ഹ ന്ദ സി സം
എ ക്കു ക ക ക - - - - -

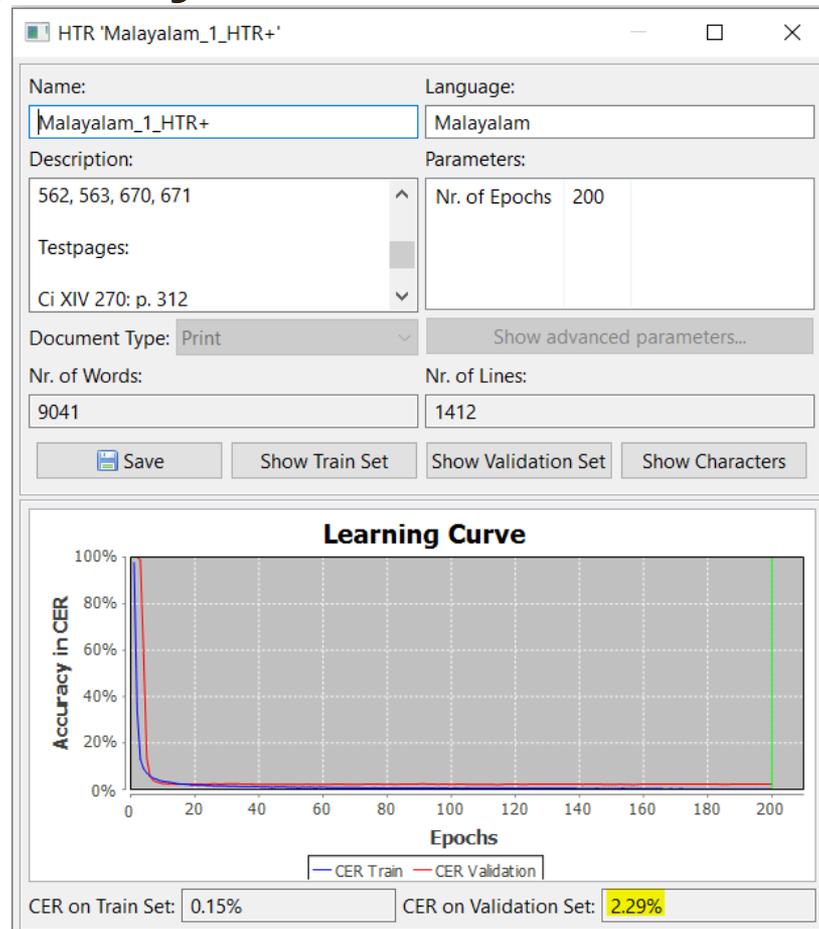
മൽവുലുളു മക...

മ ല റ റ റ ജ കൃ വ ട കെ ഹൃ ക മ റ
ക മ ക്കി റ ക്കു ട സൃ കൃ ടേ ടേ റ
കൃ പ്ക റ ഹി വി സ മ ക്കു ഹൃ വ റ ക ക
കൃ പ്ക റ റ കൃ വി റ വ ക്കു റ റ



Ergebnis Modelltraining Malayalam

- 48 S. GT aus 5 Drucken
- 2,29% CER





Veröffentlichung der GT auf GitHub

The screenshot shows the GitHub interface for the repository 'ubtue / Ground-Truth'. The repository is on the 'main' branch and has 1 branch and 0 tags. The commit history shows a commit by 'wagnerflo' on 16 Feb, with 1 commit. The repository contains several files: UAT_407_080, UAT_407_081, UAT_407_082, UAT_407_105, and README.md. The README.md file is expanded, showing the title 'Tagebücher Edwin Hennig' and a description: 'This repository contains transcriptions of selected diaries by Edwin Hennig (1882-1977), predominantly regarding the Tendaguru expedition.' The README also lists the diaries included: UAT 407/80 (1909-1910), UAT 407/81 (1910-1911), UAT 407/82 (1911), and UAT 407/105 (1897, 1899, 1901). It specifies the language as 'Mostly German, some parts in African and other European languages' and the script as 'Mainly Kurrent, some parts in Cursive'. The transcription guidelines are also mentioned.



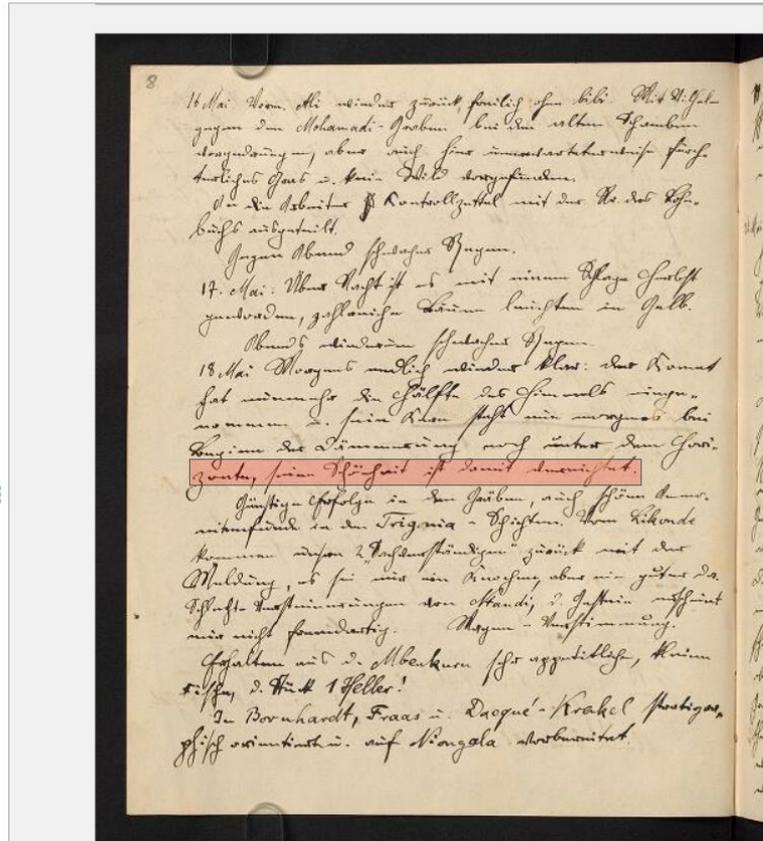
Ansicht im OpenDigi

OpenDigi > DigiUniArchiv > Edwin Hennig (1882-1977), Nachlass (p. 8)

Info Inhalt OCR-Volltext

8 ↵

16 Mai. Vorm. Ali wieder zurück, freilich ohne bibi. Mit Wilhelm ↵
gegen den Mohamadi-Graben bei den alten Schamben ↵
vorgegrungen, aber auch hier unerwarteterweise fürch, ↵
terliches Gras u. kein Wild vorgefunden. ↵
An die Arbeiter Z Kontrollzettel mit der Nr. des Lohn, ↵
buchs ausgeteilt. ↵
Gegen Abend schwacher Regen. ↵
17. Mai: Über Nacht ist es mit einem Schläge Herbst ↵
geworden, zahlreiche Bäume leuchten in Gelb. ↵
Abends wiederum schwacher Regen. ↵
18 Mai Morgens endlich wieder klar: der Komet ↵
hat nunmehr die Hälfte des Himmels einge, ↵
nommen u. sein Kern steht nun morgens bei ↵
Beginn der Dämmerung noch unter dem Hori, ↵
zonte, seine Schönheit ist damit vernichtet. ↵
Günstige Erfolge in den Gräben, auch schöne Ammo, ↵
nitenfunde in den Trigonía-Schichten. Vom Likonde ↵
kommen unsre 2 „Sachverständigen“ zurück mit der ↵
Meldung, es sei nur ein Knochen, aber ein guter da. ↵
Schlechte Versteinerungen von Ntandi, d. Gestein erscheint ↵
mir nicht fremdartig. Magen-Verstimmung. ↵
Erhalten aus d. Mbenkuru sehr appetitliche, kleine ↵
Fische, d. Stück 1 Heller! ↵
In Bornhardt, Fraas u. Daequé-Krenkel stratigra, ↵
phisch orientiert u. auf Niongala vorbereitet. ↵





Nutzeranfragen

- Orient- und Islamwissenschaft: Edition eines Briefkorpus
- Paläontologie: Transkription handschriftlicher Aufzeichnungen früherer Fachvertreter
- Biologie: handschriftliche Herbarien
- Skandinavistik: Aufbereitung einer isländischen Druckedition
- verschiedene SFBs (Geisteswissenschaften): Bearbeitung von Handschriften in unterschiedlichen Sprachen und Schriftsystemen, Druck/Handschrift gemischt
- Dokumente aus Museen und Archiven



Fazit

- Ergebnisse der automatischen Texterkennung besser als erwartet (auch heterogenes Material wie z.B. unterschiedliche Schreiber/ lange Schreibzeiträume beeinträchtigt das Ergebnis nicht wesentlich und verlangt bei entsprechender Planung nicht unbedingt einen höheren Ressourcenaufwand)
- unterschiedliche Sprachen und Schriftsysteme sind kein Problem
- für ein bestmögliches Ergebnis ist ein Modelltraining notwendig (für ein fehlerfreies Ergebnis bedarf es manueller Nachkorrektur)
- erhöhte Nachfrage und Nutzung von Handschriften, für die ein Volltext vorliegt



Ausblick für die 2. Projektphase (2021-2022)

- Bearbeitung von kleineren, heterogenen Textkorpora
- werksspezifisches Training auf Grundlage von generischen Modellen
- Evaluierung von eScriptorium / einer eigenen Transkribusinstanz in zusammen mit der UB Mannheim
- Unterstützungsangebote durch Schulungsmaterialien und persönliche Beratung



Danke.

Kontakt: Dorothee Huff

Universitätsbibliothek Tübingen
Wilhelmstraße 32, 72074 Tübingen
Telefon: +49 7071 29-72852
dorothee.huff@uni-tuebingen.de