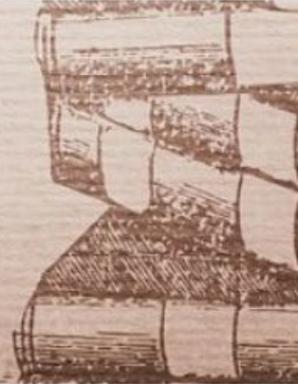


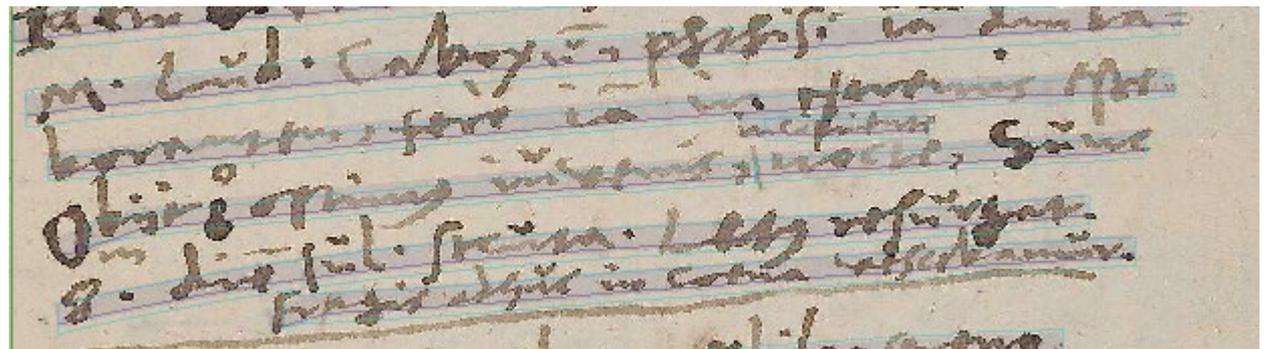


OCR-BW

Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen



Texterkennung von Handschriften mit Transkribus im Projekt OCR-BW



5.10.2021, Kristina Stöbener und Dorothee Huff



Agenda

- Projektvorstellung OCR-BW
- Vorstellung Transkribus
- Fallbeispiele
 - Modelltraining Variante 1: Tagebücher von Edwin Hennig
 - Modelltraining Variante 2: Predigtnachschriften/Tagebücher von Martin Crusius
 - Spezialmodell vs. generisches Modell: Jurist. Konsilien
- Fazit



OCR-BW (2019-2022)

- Projektpartner: UB Mannheim, UB Tübingen, gefördert durch MWK
- Projektziel: Aufbau eines Kompetenzzentrums für Volltexterschließung von handschriftlichen und gedruckten Werken
- Projekt OCR-BW unterstützt Archive, wissenschaftliche Bibliotheken und andere Institutionen in Baden-Württemberg bei der Anwendung von automatischer Texterkennungs- und Transkriptionssoftware
- UB Tübingen: Transkription und Volltexterschließung von Handschriften und Inkunabeln
- UB Mannheim: Volltexterkennung (OCR) von Druckwerken



1. Projektphase (2019-2021)

- Bearbeitung und Texterkennung für verschiedene Textcorpora
 - Tagebücher des Geologen u. Paläontologen **Edwin Hennig** (1909-1911)
 - lateinische Tagebücher (1573-1605) und griechische Predignachschriften (1563-1604) des Altphilologen **Martin Crusius**
 - ausgewählte Bände der **juristischen Konsilien** (1602-1879) und Senatsprotokolle (1524-1912)
 - Handschriften und Drucke in Malayalam
- Unterstützung von Projekten und Wissenschaftlern beim Umgang mit Transkribus (Schulungen, Workshops)



2. Projektphase (2021-2022)

- Bearbeitung von kleineren, heterogenen Textkorpora
 - mittelalterliche Handschriften
 - vermischtes, loses Schriftgut (auch Maschinschrift, 19./20. Jh.)
 - Inkunabeln
- werksspezifisches Training auf Grundlage von generischen Modellen
- Evaluierung von eScriptorium / einer eigenen Transkribusinstanz zusammen mit der UB Mannheim
- Unterstützungsangebote durch Schulungsmaterialien und persönliche Beratung + Schulungen und Workshops



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) und über 50 000 registrierte Nutzer
- Desktop-Version (Expert Client) und Browser-Version (Transkribus Lite)
- Gebühren für die automatische Texterkennung, die übrigen Funktionen sind kostenlos nutzbar



Transkribus v1.17.0 (22_09_2021_15:04). Loaded doc: Mh466-6, ID: 269164, Page 83, file: Mh466-6_083.jpg [Image Meta Info: (Resolution:399.0, w'h: 2466 * 3242)] [current line: w'h: 1754 * 203]

Server Overview Layout Metadata **Tools**

Layout Analysis

Method: CITab Advanced Configure...

Current page

Pages (863): 1-863

Document Selection

Mh466-6 (269164)

Current collection Choose docs...

Find Text Regions Empty pages only

Find Lines in Text Regions

Run

Text Recognition

Method: HTR (CITab HTR+ & PyLaia)

Models... Train...

Run

Compute Accuracy...

Reference: (Correct Text) 26.10.20 09:34:25 - handschriften@ub.uni-tuebingen.de - Ground Truth Use current

Hypothesis: (HTR Text) 26.10.20 09:34:25 - handschriften@ub.uni-tuebingen.de - Ground Truth Use current

Compare Text Versions...

Compare...

Compare Samples...

Other Tools

P2PaLA...

Text2Image...

Document Understanding...

Current page Pages (86): 1-863

Add Baselines to Polygons

Add Polygons to Baselines

1-11 hætten4, Pirates, 4

1-12 7-affines, 7c. 4

1-13 Deß herrn Crufij schreiben, würdt gwiß= 4

1-14 lich ad man9 Principis Transylua= 4

1-15 niæ, fürderlich uff der poft (dahin 4

1-16 ez auß deß Königs Veldcantzlei, 4

1-17 uff herrn von Tieffenbachs Rhat, 4

1-18 geben.) kömen. und mir wider 4

1-19 antwort wærdem. Diefelbige will 4



Vorgesehene Quellen

- Tagebücher des Geologen u. Paläontologen Edwin Hennig (1909-1911)
- lateinische Tagebücher (1573-1605) und griechische Predigt-nachschriften (1563-1604) des Altphilologen Martin Crusius
- ausgewählte Bände der juristische Konsilien (1602-1879) und Senatsprotokolle (1524-1912)
- mittelalterliche Handschriften
- vermischtes, loses Schriftgut (auch Maschinenschrift, 19./20. Jh.)
- Handschriften und Drucke in Malayalam
- Inkunabeln

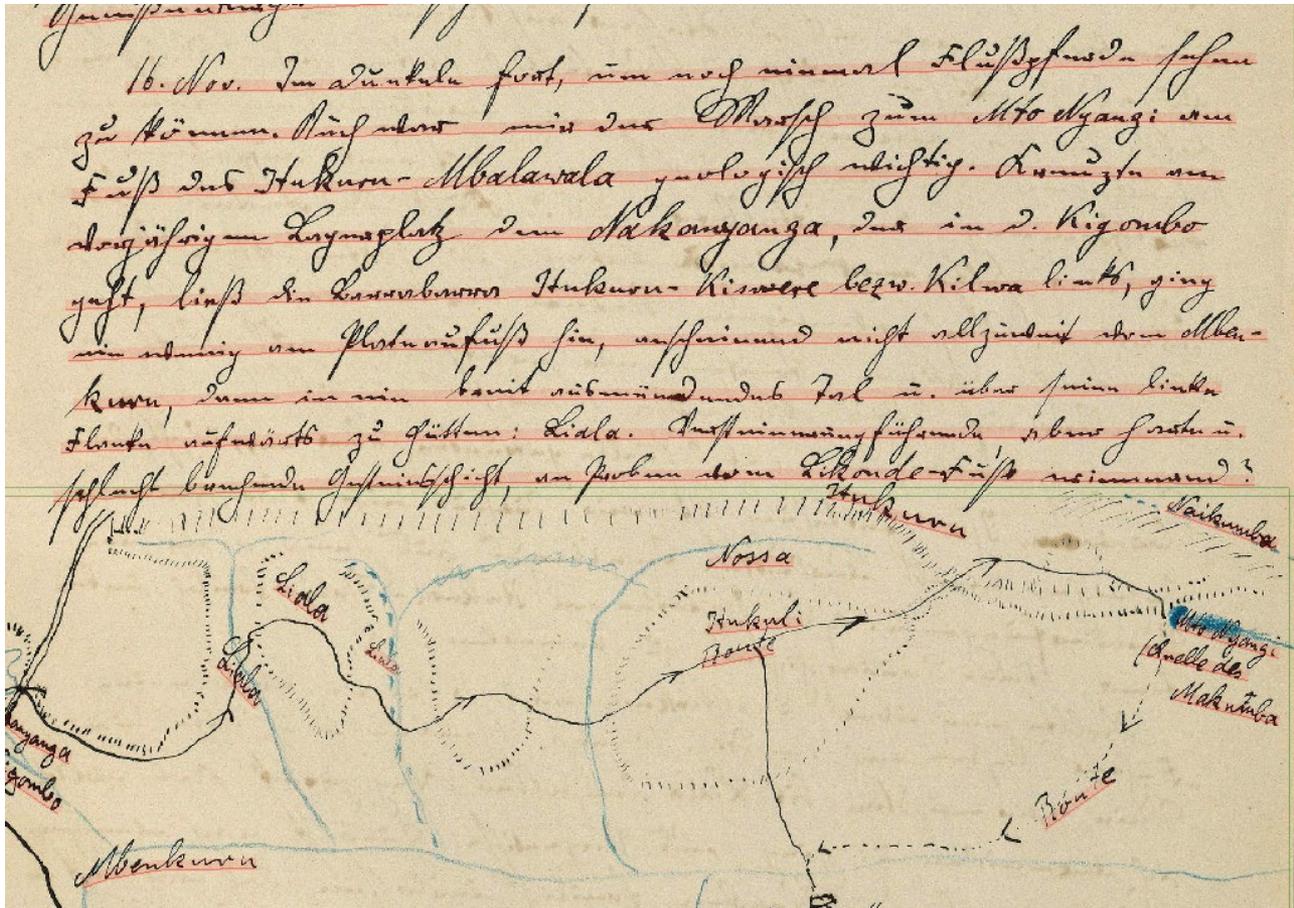


Aufgabe

- 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 ~~nüchicht-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.

 betete. 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
Flroaecears 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- orbemababiiegenauberer- 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- Naerrihmrrkoriele- 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- Neerraensderndee- 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

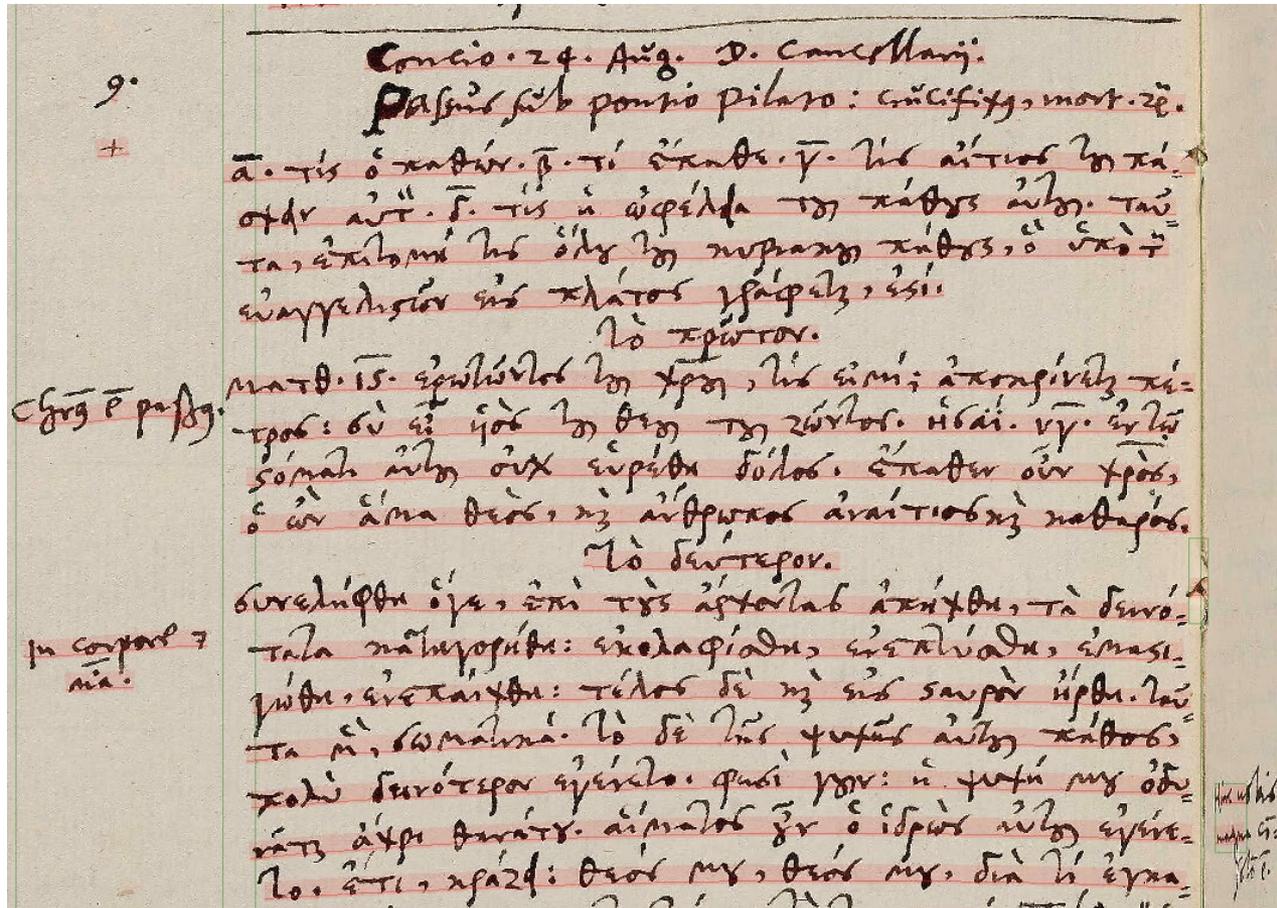
M6: CER 4.53%, WER 18.82%

(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
andern 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~~



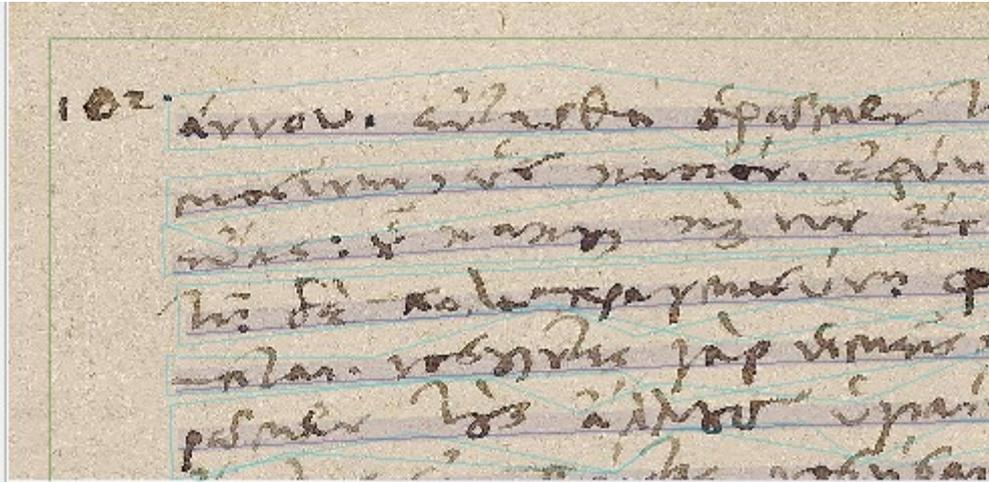
Griech. Predignachschrift, Mb 19-4, (1563-70)





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)



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= ↵



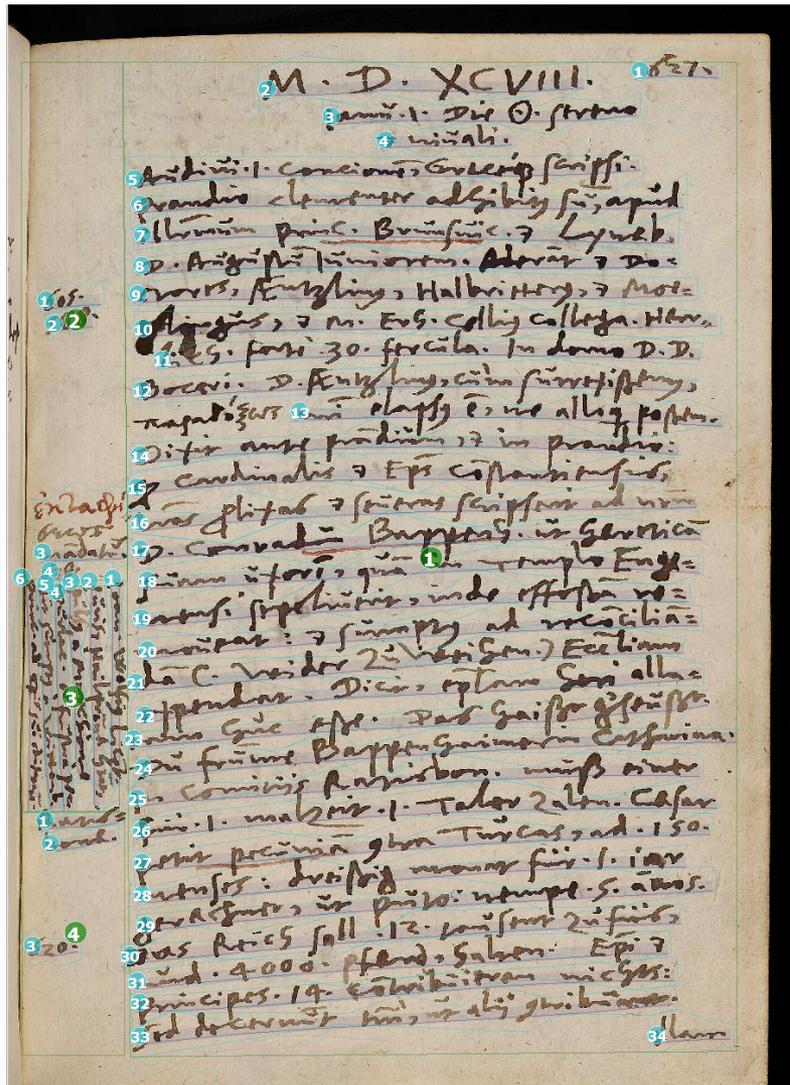
Problematik Transkriptionsrichtlinien

- Stand: bisher keine Normierung bei der Transkription von Handschriften
- Maxime: möglichst dokumentnah transkribieren
- Orientierung an OCR-D Guidelines → Level 2
- Praktikabilität (nur Unicode kein PUA/MUFI)
- Beratung durch Fachvertreter



Beispiel lateinische Tagebücher von Martin Crusius

	Auflösung	Transkription	PUA/MUFI
	et	7	
	conquestus	gqueft ⁹	
	postque	poftq̄z	

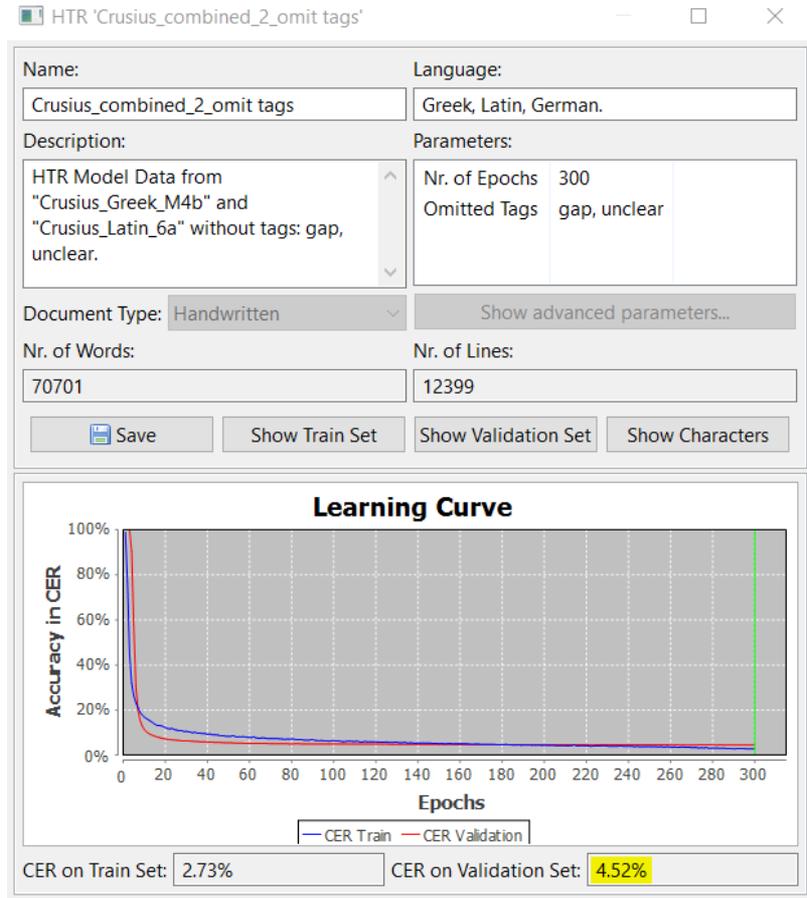


- 1-1 627. ↵
- 1-2 M. D. XCVIII. ↵
- 1-3 Ianu. 1. Die O. fereno ↵
- 1-4 7 nuali. ↵
- 1-5 Audiui. 1. concionē. Græceq̄. scripfi. ↵
- 1-6 Prandio clementer adhibi⁹. fū. apud ↵
- 1-7 illr̄mum. Princ. Brunfuic. 7 Lynæb. ↵
- 1-8 D. Auguftū. Iuniorum. Aderat. 7 Do = ↵
- 1-9 ctores. Æntzlin⁹. Halbritter⁹. 7 Moe = ↵
- 1-10 glingus. 7 M. Erh. Celli⁹. collega. Herr = ↵
- 1-11 Læh. Fortè. 30. fercula. In domo D. D. ↵
- 1-12 Boceri. D. Æntzim⁹. cūm. furrexifsem⁹. ↵
- 1-13 m̄i. elap⁹. ē. ne. alloq. pofsem. ↵
- 1-14 Dixit. ante. prædiam. 7 in. prandio. ↵
- 1-15 q̄. Cardinalis. 7 Eps. cōstantienfis. ↵
- 1-16 Irās. plixas. 7 feueras. fcripferit. ad. n̄m. ↵
- 1-17 D. conradui. Bappenij. ut. hæreticā. ↵
- 1-18 fuum. uxore. quā. in. Templo. Enge = ↵
- 1-19 nenfi. fepeliuerit. inde. effofsa. re = ↵
- 1-20 moueat. 7 finupt⁹. ad. recōciliā = ↵
- 1-21 dā. (.wider. zuweihen.) Eccliam. ↵
- 1-22 expendat. Dicit. eptam. hero. alla = ↵
- 1-23 lam. huc. else. Das. haifst. gfeufst. ↵
- 1-24 Du. frume. Bappenhaimerin. Cathaina. ↵
- 1-25 In. comitijs. Ratisbon. muß. einer. ↵
- 1-26 fur. 1. malzeit. 1. Taler. zalen. Cæfar. ↵
- 1-27 petit. pecuniā. q̄tra. Turcas. ad. 150. ↵



Kombinationsmodell Crusius

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

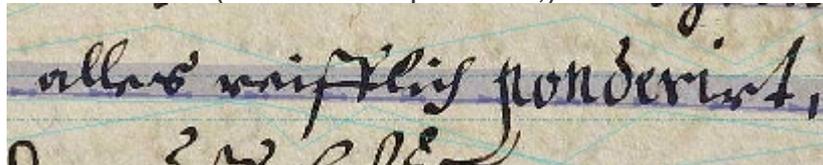




Juristische Konsilien

- 269 Bände (1602-1883)
- bearbeitet wurden Bd. 84/13 (1659-1661) und 84/14 (1659-1661)
- Sprache: Deutsch, Latein
- Schrift: deutsche Kurrentschrift, lateinische Schreibschrift

(alles reifflich ponderirt,)





Herangehensweise: Erstellung der GT

- Transkription auf Grundlage eines Public Models für Kurrentschrift

Documents **HTR Model Data**

All ▾ All engines ▾

1-9 / 9 1 1

Name	Size	Curator	HTR...	nrOfWor...	Date
> Transkribus German Kurrent M2		guenter.hackl@...	32524	3209689	27.04.21
> StAZH_RRB_German_Kurrent_XIX		tobias.hodel@...	30919	26026908	14.03.21
> Transkribus German Kurrent M2		guenter	29820	3209690	28.01.21
> Land registers (Verfachbücher) Tyrol, 1750-1800		p.andorfer@g...	28866	129152	21.12.20
> German_Kurrent_17th-18th		alverman@uni-...	27457	1839841	10.11.20
> VB_1750_1800_M1Base		p.andorfer@g...	26269	106794	24.09.20
> German_Kurrent_XIX_pylaia		tobias.hodel@...	26068	5100439	11.09.20
> German_Kurrent_XIX_M2		tobias.hodel@...	23847	5100439	15.05.20
> German_Kurrent_XVI-XVIII_M1		tobias.hodel@...	19584	1579208	16.12.19
> Acta 17 (extended)		alverman@uni-...	32677	1632534	04.05.21
> Acta_17 PyLaia		alverman@uni-...	27337	594628	04.11.20
> Acta_17 HTR+		alverman@uni-...	26284	594628	25.09.20



Modellentwicklung: 1. Modell (26 S. GT)

HTR 'Juristische_Konsilien_Tuebingen_M1a'

Name: Juristische_Konsilien_Tuebingen_M1a Language: German, Latin

Description: First model based on UAT 84/13, one scribe, p. 25-39, 102-112 (=26 p.), automatic selection of validation set 10% (p. 34-35, 107)

Parameters: Nr. of Epochs 200

Document Type: Handwritten Show advanced parameters...

Nr. of Words: 4029 Nr. of Lines: 610

Save Show Train Set Show Validation Set Show Characters

Learning Curve

CER on Train Set: 0.12% CER on Validation Set: 6.21%

HTR 'Juristische_Konsilien_Tuebingen_M1a_basemodel'

Name: Juristische_Konsilien_Tuebingen_M1a_basem Language: German, Latin

Description: First model based on UAT 84/13, one scribe, p. 25-39, 102-112 (=26 p.), automatic selection of validation set 10% (p. 34-35, 107),

Parameters: Nr. of Epochs 200
HTR Base Model ID 26284
HTR Base Model Name Acta_17_HTR+

Document Type: Handwritten Show advanced parameters...

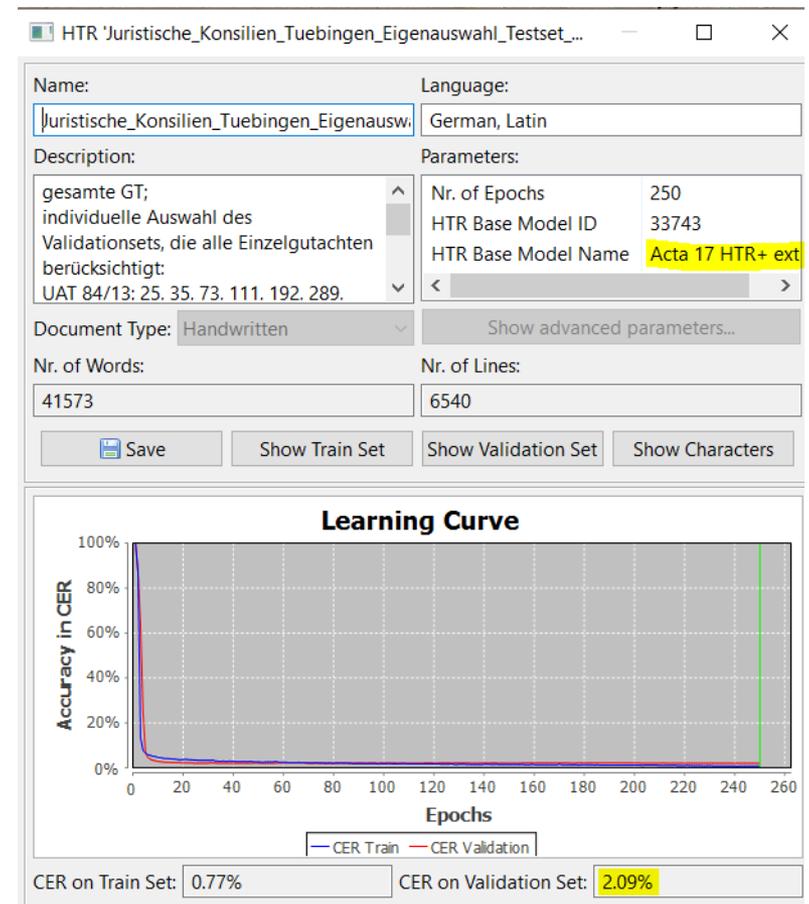
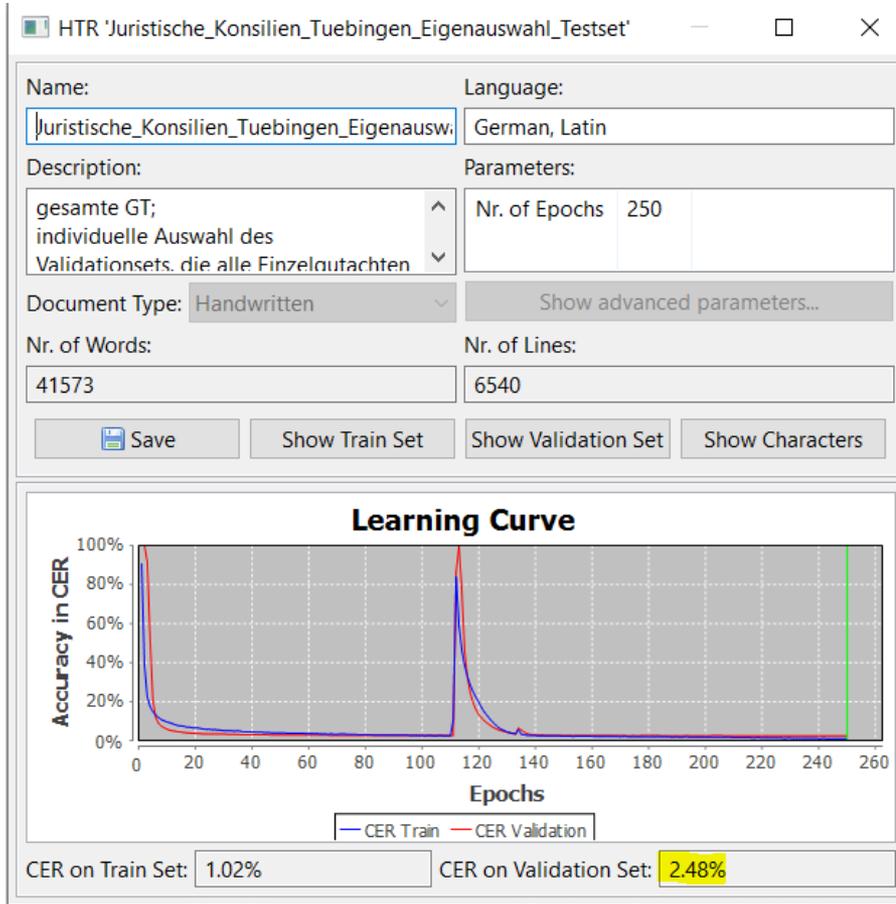
Nr. of Words: 4029 Nr. of Lines: 610

Save Show Train Set Show Validation Set Show Characters

Learning Curve

CER on Train Set: 0.07% CER on Validation Set: 4.72%

Modellentwicklung: Endergebnis (223 S. GT)



Modell „Juristische_Konsilien_Tuebingen_Eigenauswahl_Testset_Basemodel2“

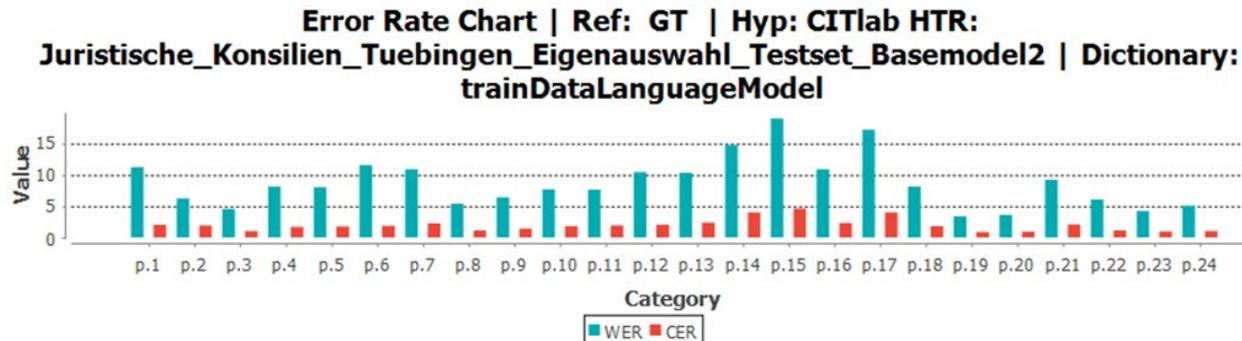
- durchschnittliche CER mit LM: 1,95%

Advanced Statistics

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Overall	8.59 %	1.95 %	86.08 %	96.88 %	0.931	0.934	0.931

Page	WER	CER	Word Acc	Char Acc	Bag Tokens ...	BT Recall	BT F1-Score
Page 1	11.22 %	2.06 %	82.81 %	97.0 %	0.9347	0.9073	0.92079207...
Page 2	6.23 %	1.92 %	90.75 %	96.97 %	0.96829999...	0.9506	0.95937090...
Page 3	4.59 %	1.01 %	92.89 %	98.39 %	0.9769	0.96790000...	0.97235023...
Page 4	8.17 %	1.68 %	86.81 %	97.57 %	0.9439	0.9346	0.93924466...
Page 5	8.03 %	1.77 %	87.59 %	97.3 %	0.93599999...	0.93980000...	0.93787575...

Compare Text Versions for Page ..





0,88% CER

Cum tamen Reus, maxime in
 gorie, perspicue & rotunde
 nis membra respondere debeat
 fieri possit, quidnam ab actore
 tenus atq; in quibus membris on
 gere debeat; adeo, ut, si reus
 spondere nolit, & in hac contu
 ad veritatem eruendam tortu
 Carpz. prax. crim. part. 3. q. 1
 Ubicunq; autem clara & rotunda
 ibi non sufficiunt talia verba d
 in specie de litis contestatione, q

459.

Cum tamen Reus, **maximè maxime** in criminalibus, cate,
 gorie, **perspicie perspicue** & rotunde ad singula accusatio,
 nis membra respondere debeat, ita ut Judex certus
 fieri possit, quidnam ab actore probandum sit, & qua,
 tenus atq; in quibus membris onus probandi ei injun,
 gere debeat; adeo, ut, si reus categorice & **clare clarè** re,,
 spondere nolit, & in hac contumacia persistat, ipse
 ad veritatem eruendam torturæ subjici **posfit posfit.**

Carpz. prax. crim. part. 3. q. 106. n. 72. & seqq.

Ubicunq; autem clara & rotunda **requiritur, responsio,, requiritur responsio,**

ibi non sufficiunt talia verba dubia & ambigua; quod
 in specie de litis contestatione, quod sc: illa per talia
 verba ambigua fieri nequeat, tradit

Wesemb. in Cod. tit. de litis contestat. n. 6. in f. per
 l. si **debitor. debitor,** ff. de heredit. petit.

Unde lis non censetur contestata, si reus respon,,
 derit: credo **ine-me** non teneri; vel credo insti,,
 tutam actionem non esse veram.

Accurf. in l. un. C. de lit. contestat.

Quod maximè Quod maxime in criminalibus procedit, et in
 illis talem responsionem litis contestationem non
 inducere, docet

Obrecht. tr. de litis contest. cap. 7. n. 6. & 7. ubi
 subjicit in criminalibus non sufficere credulita,,
 tis responsionem, sed veritatis exigi. Carpzov.

in Jurisp. forens. part. 1. const. 16. def. 6. Umm. **ad-ad**
 procesf. disp. 12. th. 4. n. 14. Brunnemann. de procesf.
 cap. 14. n. 15.

Jdem tradunt Dd. si reus dicat: neq; nego, neq;



4,64%

112.
übel tractirt und gehalten, d3 nicht
ein geringer argwohn sich ereigen
wollen, als ob folches vier Jähriges
Knäblein, Knäblein, durch übele fractation-tractation
und haltung, sein leben-Leben eingebueßt
hette, und verstorben were, were.
Wir Wie dann hiervon gleichsam fama publica publica,
und ein allgemein geschrey entständen-entstanden
ist, auch die verhörte zeug-zeug uno ore
bekundtschafften thuen. Gestaldt-dae Gestaldtsame, in
specie specie, Michel Klenckh sagt: d3 Kindt
puta (puta d3 verstorbene Knäblein) habe nit
adlein-allein
nicht
igung „gnug zu bfen-Eben bekoimen mögen,
fond'n feye auch über die maßen hart
Ghalten, Gehalten, und immer-immer zu von des Jacob
Brunen weib gehaft, geschlagen, geschleiffet,
und im-un Christlich tractirt worden. Mit
welchem zeug auf-auch
Durchhaß-durchhaß
übereinstimmen, Michel
dung-Jung der 13. wie auch Melchior Jung
deri4-der 14. Item Hanß Georg Schon-Schön der 18
Zeug, welche alle, nebenland'n mehr,, neben and'n mehr,
einhelliglich, von folcher des Kinds
ußgestandtener böfen tractation
attestiren thuen. Margaretha
hauß wollen Schneids-Tochter Hanß Wolfen Schneid's tochter die
25 gengin-zeugin zeigt an, Mann habe d3 Kind
aufgehungert, aufgehüngert, biß weihn ihme Brot-brot
eingebrocket, und darüber nur
bloß waßer geschittet, welches d3 Kind
nicht genüßen Können, fondern wid'
von sich geben müßen. dahero vonhard-müßen. Dahero Leonhard
se vor-Ruopp vermeinen undt mittmaßen-muttmaßen
thutt, d3 Kindt feye im Leiblein-leiblein undt
mäglin, daureh-durch gegebene harte Steiß-Speiß
undt tränkch, tranckh, ganzt verderbt worden,

112.
übel tractirt und gehalten, d3 nicht
ein geringer argwohn sich ereigen
wollen, als ob folches vier Jähriges
Knäblein, Knäblein, durch übele fractation-tractation
und haltung, sein leben-Leben eingebueßt
hette, und verstorben were, were.
Wir Wie dann hiervon gleichsam fama publica publica,
und ein allgemein geschrey entständen-entstanden
ist, auch die verhörte zeug-zeug uno ore
bekundtschafften thuen. Gestaldt-dae Gestaldtsame, in
specie specie, Michel Klenckh sagt: d3 Kindt
puta (puta d3 verstorbene Knäblein) habe nit
adlein-allein
nicht
igung „gnug zu bfen-Eben bekoimen mögen,
fond'n feye auch über die maßen hart
Ghalten, Gehalten, und immer-immer zu von des Jacob
Brunen weib gehaft, geschlagen, geschleiffet,
und im-un Christlich tractirt worden. Mit
welchem zeug auf-auch
Durchhaß-durchhaß
übereinstimmen, Michel
dung-Jung der 13. wie auch Melchior Jung
deri4-der 14. Item Hanß Georg Schon-Schön der 18
Zeug, welche alle, nebenland'n mehr,, neben and'n mehr,
einhelliglich, von folcher des Kinds
ußgestandtener böfen tractation
attestiren thuen. Margaretha
hauß wollen Schneids-Tochter Hanß Wolfen Schneid's tochter die
25 gengin-zeugin zeigt an, Mann habe d3 Kind
aufgehungert, aufgehüngert, biß weihn ihme Brot-brot
eingebrocket, und darüber nur
bloß waßer geschittet, welches d3 Kind
nicht genüßen Können, fondern wid'
von sich geben müßen. dahero vonhard-müßen. Dahero Leonhard
se vor-Ruopp vermeinen undt mittmaßen-muttmaßen
thutt, d3 Kindt feye im Leiblein-leiblein undt
mäglin, daureh-durch gegebene harte Steiß-Speiß
undt tränkch, tranckh, ganzt verderbt worden,



2,13%

Cum tamen Reus, maxime in
gorice, perspicue & rotunde
nis membra respondere debeat
fieri possit, quidnam ab actore
tenus atq; in quibus membris on
gere debeat; adeo, ut, si reus
spondere nolit, & in hac contu
ad veritatem eruendam tortu
Carpz. prax. crim. part. 3. q.
Ubicunq; autem clara & rotunda
ibi non sufficiunt talia verba
in specie de litis contestatione,

45 D. 459.

Cum tamen Reus Reus, maxime in criminalibus, cate
gorice, perspicue perspicue & rotunde ad singula accusatio
nis membra respondere debeat, ita ut Judex certus
fieri possit, quidnam quidnam ab actore probandum sit, & qua
tenus atque in quibus membris onus probandi et ei injun
gere debeat, debeat; adeo, ut, si reus categorice & clare re
spondere nolit, & in hac contumacia persistat, ipse
ad veritatem eruendam torturae subijci possit possit.

Carpz. prax. crim. part. 3. q. 106. n. 22. et 72. & seqq.

Ubicunque autem clara & rotunda requiritur responsio,
ibi non sufficiunt talia verba dubia & ambigua, ambigua; quod
in specie de litis contestatione, quod se. sc. illa per talia
verba ambigua fieri nequeat, tradit

Wesemb. in Cod. tit. de litis contestat. n. 6. in f. per
I. l. si debitor. debitor, ff. de heredit. petit.

Unde lis non censetur contestata, si reus respon
derit. Credo derit: credo me non teneri teneri; vel credo insti
tutam actionem non esse veram veram.

Accur. Accurs. in I. l. un. C. de lit. contestat.

Quod maxime in criminalibus procedit, et in
illis talem responsionem litis contestationem non
inducere, docet

Obrecht. tr. de litis contest. cap. 7. n. 6. et 7. ubi
subjicit in criminalibus non sufficere credulita
tis responsionem, sed veritatis exigi exigi. Carpzov.

in Jurisp. forens. part. 1. const. 6. 16. def. 6. Umm. ad
process. disp. 12. th. 4. n. 14. Brunemann. de process.
cap. 14. n. 15.

Idem Idem tradunt Dd. si reus dicat. dicat: neque nego, nequi neque



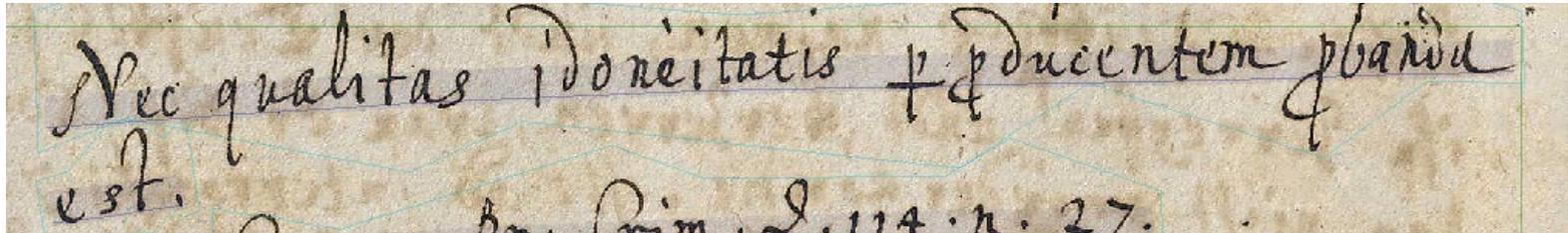
5,90%

112.
 übel tractirt und gefaltet, es nicht
 ein geringen gering argwohn sich ereigen
 wollen, als ob solches vier Jähriges
 Knäblein, durch übele tractation
 und faltung, sein Leben eingebueßt
 hette, und verstorben were etc. were.
 Wie dann hier von hiervon gleichsam fama publica publica,
 und ein allgemein geschrey entstanden
 ist, auch die verhörte Zeugen zeugen uno ore
 bekundtschafften thuen. Gestaltt same, Gestaltsame, in
 specie specie, Michel Kleuth-Klenckh sagt: das Kindt
 puta (puta das verstorbene Knäblein) habe nit
 allein
 nicht
 genug zu Eßen bekommen bekommen mögen,
 sondern wege seye auch über die maßen hart hart
 Gehalten, und immer zu von des Jacob
 Brunen weib gehaft, gehast, geschlagen, geschleiffet,
 und im un Christlich tractirt worden, worden. Mit
 welchen Zeugen welchem zeugen auch
 Durchaus-durchauß
 übereinstimmen, übereinstimmen, Michel
 Jung der 13. Wie wie auch Melchior Jung
 den der 14. Item Item Hanß Georg Schon-Schön der 18
 Zeuge, Zeug, welche alle, neben andern mehr= andern mehr,
 einhelliglich, von solcher des Kinds
 außgestandener außgestandener bösen tractation
 attestiren thuen. Margaretha
 hauß wolten-Schneids-Hanß Wolfen Schneiders tochter die
 25 Zeugin zeugin zeigt an, Mann habe das Kin-Kind
 außgehüngert, außgehüngert, biß weiln ihme Grot-brot
 eingebrochet, eingebrochet, und darüber nur
 bloß waßer geschittet, welches das Kind
 nicht genußen können, genußen Können, sondern wider
 von sich geben mußten. Dahero vorhand-Leonhard
 Quorg-Ruopp vermeinen undt mittmaßen muttmaßen
 thut, thut, das Kindt stege-seye im leiblein undt
 mäglin, durch gegebene harte Speiß
 undt frauen, gantz Verderbt-tranckh, gantz verderbt worden,

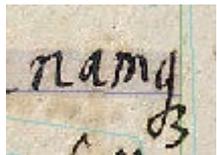
112.
 übel tractirt und gehalten. Gehalten, das nicht
 ein geringen-gering argwohn sich ereigen
 wollen, als ob solches vier Jähriges
 venablen-Knäblein, durch übele tractation
 undt haltung, kein leben sein Leben eingebueßt
 hette, und verstorben were etc. were.
 Wie dann hier von hiervon gleichsam fama publica publica,
 und ein allgemein geschrey entstanden
 ist, auch die verhörte Zeugen zeugen uno ore
 bekundtschafften thuen. Gestaltt same, Gestaltsame, in
 specie specie, Michel Kleuth-Klenckh sagt: das Kindt
 puta (puta das verstorbene Knäblein) habe nit
 allein
 nicht
 genug zu Eßen bekommen bekommen mögen,
 sondern wege seye auch über die maßen hart hart
 Gehalten, und immer zu von des Jacob
 Brunen weib gehaft, gehast, geschlagen, geschleiffet,
 und im un Christlich tractirt worden, worden. Mit
 welchen Zeugen welchem zeugen auch
 Durchaus-durchauß
 übereinstimmen, übereinstimmen, Michel
 Jung der 13. Wie wie auch Melchior Jung
 den der 14. Item Item Hanß Georg Schon-Schön der 18
 Zeuge, Zeug, welche alle, neben andern mehr= andern mehr,
 einhelliglich, von solcher des Kinds
 außgestandener außgestandener bösen tractation
 attestiren thuen. Margaretha
 hauß wolten-Schneids-Hanß Wolfen Schneiders tochter die
 25 Zeugin zeugin zeigt an, Mann habe das Kin-Kind
 außgehüngert, außgehüngert, biß weiln ihme Grot-brot
 eingebrochet, eingebrochet, und darüber nur
 bloß waßer geschittet, welches das Kind
 nicht genußen können, genußen Können, sondern wider
 von sich geben mußten. Dahero vorhand-Leonhard
 Quorg-Ruopp vermeinen undt mittmaßen muttmaßen
 thut, thut, das Kindt stege-seye im leiblein undt
 mäglin, durch gegebene harte Speiß
 undt frauen, gantz Verderbt-tranckh, gantz verderbt worden,



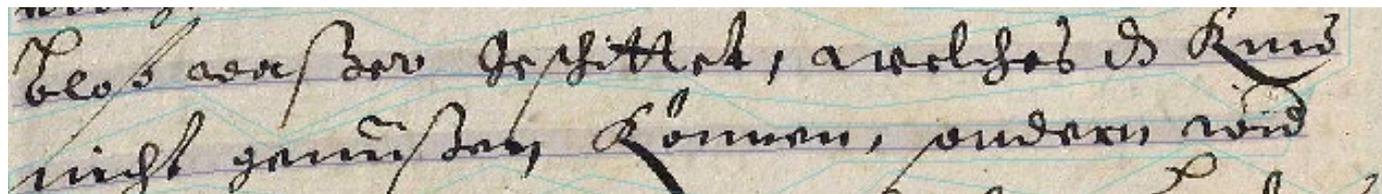
Auswirkung der Transkriptionsrichtlinien



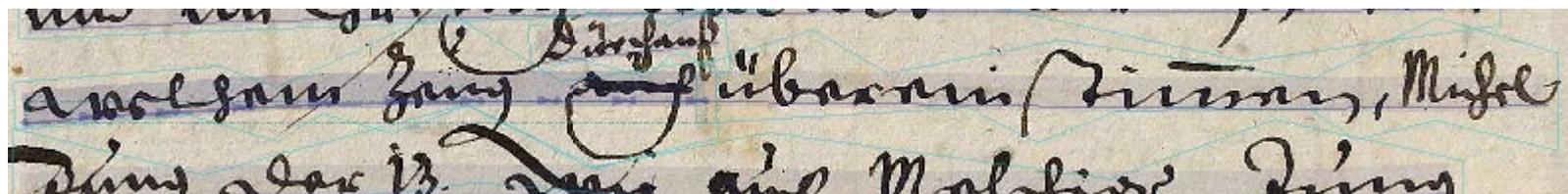
- **Jur_Kons_Tue:** Nec qualitas idoneitatis p̄ p̄ducentem p̄bandu
- **Acta_17:** Nec qualitas idoneitatis per producentem p̄banda



- **Jur_Kons_Tue:** namq̄
- **Acta_17:** namque



- **Jur_Kons_Tue:** bloß waſſer geſchittet, welches **d3** Kind nicht genüßen können, ſondern wid
- **Acta_17:** bloß waſſer geſchittet, welches **das** Kind nicht genußen können, ſondern wid**er**



- **Jur_Kons_Tue:** welchem zeug **u** auf Durchanß **übereintimen**, Michel
- **Acta_17:** welchen Zeugen **en** auch Durchaus übereinstim**en**, Michel



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, wobei sich oftmals auch mit generischen Modellen gute Ergebnisse erzielen lassen (für ein fehlerfreies Ergebnis bedarf es manueller Nachkorrektur)



Danke.

Kontakt: Dorothee Huff

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