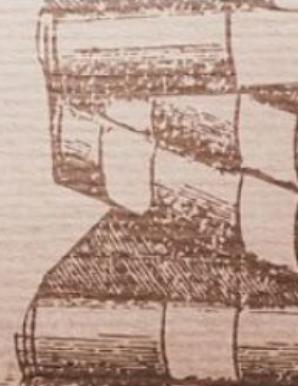


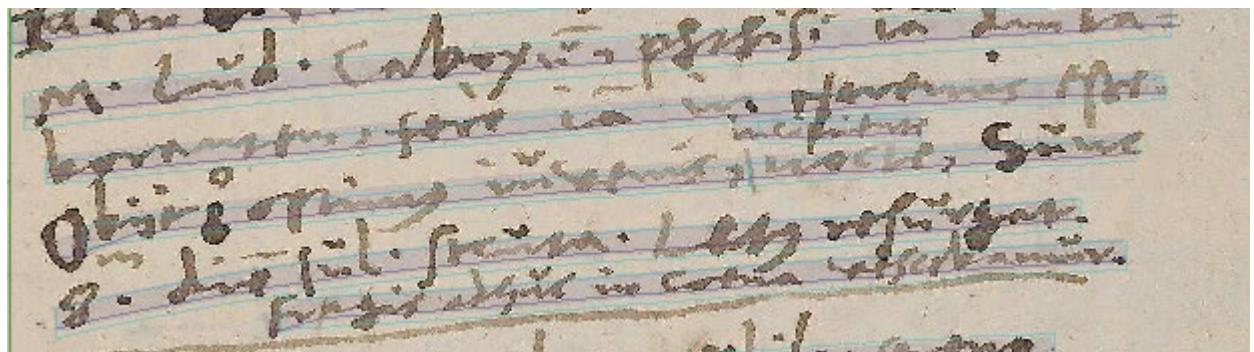


OCR-BW

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: Predigt-Nachschriften/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 Predigt-Nachschriften (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 Maschinenschrift, 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: CITLab Advanced Configure...

Current page
Pages (863): 1-863

Document Selection
Mh466-6 (269164)
Current collection Choose docs...
Find Text Regions Find Lines in Text Regions Run

Text Recognition
Method: HTR (CITLab HTR+ & PyLaia)
Models... Train... Run...

Compute Accuracy...
Reference: 26.10.20 09:34:25 - handschriften@ub.uni-tuebingen.de - Ground Truth Use current
(Correct Text)
Hypothesis: 26.10.20 09:34:25 - handschriften@ub.uni-tuebingen.de - Ground Truth Use current
(HTR Text)
Compare Text Versions... Compare... Compare Samples...
Other Tools P2PaLA... Text2Image... Document Understanding...

Current page Pages (863): 1-863 Add Baselines to Polygons Add Polygons to Baselines

Annotations:

- 1-2 affines, 7c. ↗
- 1-13 Deß-herrn-Crufij-schreiben, würdt-gwiß= ↗
- 1-14 lich-ad-man⁹ Principis-Tranſylua= ↗
- 1-15 niæ,fürderlich-uff-der-poſt-(dahin) ↗
- 1-16 ez-auß-deß-Königs-Veldcantzlei, ↗
- 1-17 uff-herrn-von-Tieffenbachs-Rhat, ↗
- 1-18 geben.)-kömen.-und-mir-wider ↗
- 1-19 antwort-wärden. Dieſelbige-will ↗



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

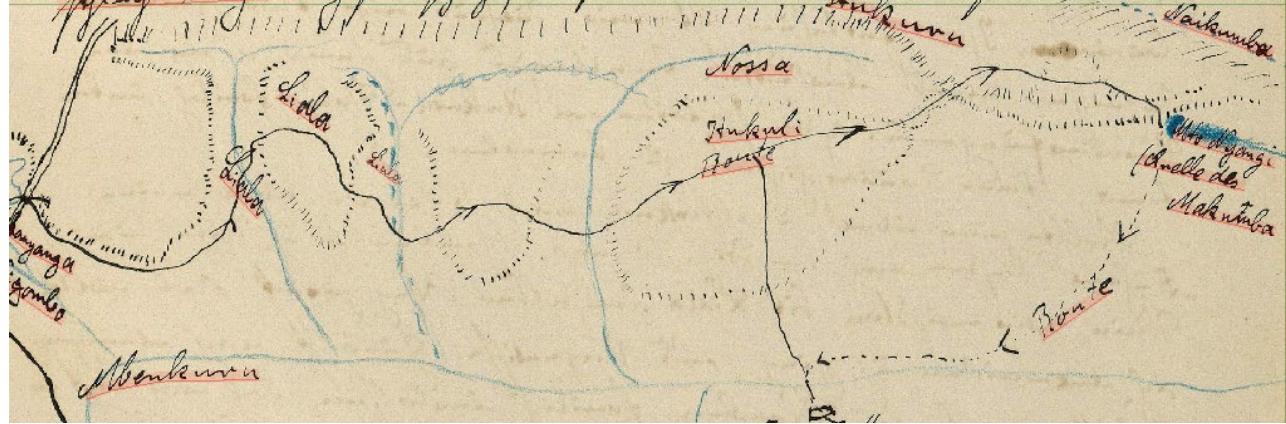


Aufgabe

- 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)

16. Nov. Im Abendklu fand, von mir wieder Fließpferde gesen
ge können. Auf dem mit dem Massiff zum Mt. Elangai von
Fopp das Hukuru-Mbalawala zuverlässig reitfähig. Kommen nur
dort jenseit der Lymaghat dem Nakanyanga, da im D. Kigombo
gut, ließ ein Passerbruder Hukuru-Kisere bzw. Kilwa links, ging
nun aber vom Flussniveau fluss hin, aufgerichtet nicht allzudeutlich ob der
Kewu, wenn er nun leicht absteigt und die Tiere in übern jenen linken
Flusku wiederkommt zu füttern: Siala. Die Pferde aufgerichtet oben großer
gestalt bengende Geprangfößig, um Pferden den Rückende-füß reizend?





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/LanguageModel		Comb_Deutsch_Latein_XX_M2	German_Kurrent_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 Trainingsdatensatz	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



8

Löffel. Durch Lüfte und ipsum selbst in möglichst viele Richtungen, also auf
ging ich mit schneller Landungslösung bis zu den überwiegend aufgeweichten Fliegen
(ca. 2/3 der Lüfte pflügen z.T. mit Schaufelköpfen!), in jenseits der Richtungsflug.
Ich ging unmittelbar nach einer 2. Landung dort, nicht mehrum, aber ich war I.
Gleichwohl sollte man nun feststellen, dass Pärchen Gruppe auf Sichtlinie 25 m.
nach jedem anderen, eben das nicht mit zwei Gruppen d. 2. Gruppe geflogen
werden. Dies fliegt ihm natürlich nur für abwechselnd auf und zu kommen
kommt. Die Säulen sind diese nicht. Da dies nicht möglich ist,
z. B. wenn zu einem anderen wollen ich in einer unorganisierten Gruppe (abfahrt
Halt) auf die für zu geflüchtet werden. Wenn sie dann gegenfliegen,
dann wird absichtlich eine gezielte Richtung einhalten, sobald sie fliegen
nicht auf d. anderen Seite d. Gruppe hingehen. Aber wenn wir das
nicht unvermeidlich, zuerst sind sie auf dem ersten 5 m. abge-
blieben und dann weiter in. so sehr nicht wie ein beweitete
Fliegen und mit Löffel kommt. Dies ging am Ende von Gruppe nicht
in d. Richtung, wo ich selbst zu fahren (als die Halt-abfahrt den Fliegen, die
hier - unvermeidlich beim Zufahren auf der Straße d. Motorrad i. Z. Kehle).



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 (a-(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 starken-starken Herde auf vielleicht 25 m. nate nahe gekommen, ohne daß auch nur das Geringste v. d. Tieren sichtbar wurde. Wir flüchteten eilist, eilist, 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 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., stärmender 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. Marua! Mavudy 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 üb. 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 seltsam einige seltsame Erfahrungen reicher. Die Leuthen Leutchen hier in diesem verlorenen Talkessel kamen mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.
4. Aug. Rückkehr nach 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 eilist, eilist, 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. Mavudy 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 üb. 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 Leutchen hier in diesem verlorenen Talkessel können kamen mir sehr viel freundlicher vor als alle andern Kilwa-Bewohner.
4. Aug. Rückkehr nach Makangaga. Morgens schöne Wanderung



beitete. Um 21 gingen wir nach Kurths
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,
ging 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 vorkam, da auch Kellner
davaren. Nachher gingen wir weiter und
merkten auch, dass wir garnicht beim För-
ster gewesen waren, sondern kamen erst
jetzt daran vorbei und stiegen auf einen
R... und dem einen Baum, wo Dornen waren.



M1: CER 57.69%, WER 93.73%

4-7.)

belete aaerirnachdr-beitete. Um ½1 gingen wir nach Kurth's Barlharsvne afndrrsvehrgabas Gasthaus * und aben dort sehr gut zu Nitbag. Ba ders gmmarmibeerrach Mittag. Von dort gingen wir wieder nach Nlaave nad raren ihlarhirbwir Gle-Hause und waren unschlüßig, ob wir gehenollbeneerasl. Sharbaibobenreh. hen sollten, oder nicht. Ich arbeitete noch Flroaeears vr Mrr wer rermbl. etwas und, da es nur ganz wenig regnete, KolbbienrurwBadlergeser. Lbwir. wollten wir zum Förster gehen. Als wir aa Mik ein Stück die Shaaaa Geerepraarer Chaussee gegangen waren, EineDa Bleeskeralaen dochaga fng es an stärker zu regnen, doch gingen Moraeebonne eibiolebalembar. wir weiter und es hörte bald wieder auf. Pirkehrlern MaemroenNalauruk. Wir kehrten in einem großen Restaurant IhnenPapellee Nagrrarer mit einer Kapelle ein, das mir gar nicht kbermjrrbara. Laaachiellari sehr förstermäßig vorkam, da auch Kellner NawaremNalkaoroaaamraeraaaw dawarens. Nachher gingen wir weiter und kerkiendaci daibeaarrabbi. Foru merkten auch, daß wir garnicht beim Förler. Gererer worvo, Wadera Karenerss ster gewesen waren, sondern kamen erst d. MiabDaraa orbemababiegenauberer jetzt daran vorbei und stiegen auf einen Berye mrrdura Berg, auf dem einen Turm war. Bieserae. Dieser war Aaber rereklosseruadr Neeenaden aber verschlossen und wir stiegen an den Dbrond hihab. Dordsmaarein omekya Strand hinab. Dort war ein ziemlich großer Ber, aber BrnjerPandberabesaus junger Hund, der alles aus dem Maverkeraushalae rraswankra. Wasser herausholte, was man hineinwarf, brih. Naerrihmrrkorile doch da er nicht schwimmen konnte, Donderrrdowers, aberBaraahake ging er nur so weit, als er Grund hatte, Nae, abenkeererbruaererber Waro und, als ein kleinerer Hund weiter hinaus-awarra Nararo Nedisch Naeer Aer schwamm, war er so neidisch, daß er den Mdr rbwearberawrrarerlebe andern nicht wieder herauslassen wollte balabaraikbarrerjaelba wellen und ihn nachher verjagte. Als wir weiter denilaiberakelBornke nngenaben! der Kaiser Wilhelm Brücke zugingen, sahen esenraberoillen Berenbogerra-wir einen wundervollen Regenbogen von Neodrog a Neerraainsdernde Misdroy aus ins Meer ragen, denn die Grnbwar Mederkemrraeboojerwa Sonne war wieder hervorgekommen und Kooneker ben * vorher be- zliaBessurir stiegen wir rock. 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 aben dort sehr gut zu Mittag. Von dort gingen wir wieder nach Hause und waren unschlüßig, ob wir gehen sollten, oder nicht. Ich arbeitete noch etwas und, da es nur ganz-ganz wenig regnete, wollten wir zum Förster gesen-gehen. Als wir ein Stück die Chaussee-Chaussée gegangen waren, fng es an stärker zu reehnen-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 dawarens. Nachher gingen wir weiter und merkten auch, daß wir garnicht beim Hor-Förster gewesen waren, sondern kamen erst jetzt daran vorbei 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 großer Ber, aber junger Hund, der alles aus dem Wasser herausholte, was man hinein-wart, hineinwarf, 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 heraus-sassen-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 Misdroy-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. Predigtnachschrift, Mb 19-4, (1563-70)

Concio. 24. Aug. D. Concessarij.

Passen sich ponio pilato: crucifixus, mort. 28.

τοις ὄντας. β. τι θέντε. γ. τις αἰτεῖται τινά
οὐδὲ αὐτό. δ. τις ἡ νίκησθαι τοι καθός αὐτοί. ταῦ
τα, ἀπόλονται τοι ὅλη την πρώτην καθός, εἰς οὐκεῖν
εὐαγγελιστούς τοι καθός γνήσια, εἰσιν.

τὸ πῶτον.

Cgrg P pastg. ματθ. 15. ἐργάζετο τοι χρόνος, τις εἰμί; ἀπογίνεται
τοι: εἰς τις οὐδὲ τοι διαν την γένοτο. οὐδεί. νῦν τοι
σόματιν οὐδὲ εἰπεῖται δόλος. σταύτερος οὐρ τρόπος,
εἰς τοι ἀλλα διότο, εἰς αἴρωντας ἀνάτολην γεναῖσθαι.

τὸ δευτέρον.

εὐελπίζειν εἶται, εἰτι τοις ἀγαπαῖσθαι, τὰ δενο-
τατα μάλιστα: εἰδαγίαδις, εἰσαλίαδα, εἰμα-
ριδα, εἰσταῖδα: τότος δι τοις σαφὲς ἕπειται.
τα τοι, σωματινα. τὸ δι τοις φυσεῖς αὐτοῖς καθός
τοι διερύτερος εἶναι. φαίνεται: εἰς γενήν μεγ αὖταις
νέας αἴτιος θεοῖς. αἴματος γε εἰς ιδητές μεταγενέ-
ται, εἰτι, γένεται: δεοτε μεγ, δεοτε μεγ, δια τοι εγεν-

in corpore
mia.



Griechische Predigtnachschriften von Martin Crusius

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



102. ἀννου· ἐνταῦθα ὄρῶμεν· τὴν· πολυπραγ= ↗
 μοσύνην, ὡς· κακόν· ἐφύη· αὕτη· ἐκ· τῆς ↗
 εὔας· οὐ· κακοῦ· καὶ· νῦν· ἔτι· ἀπολαύομεν. ↗
 τῇ· δὲ· πολυπραγμοσύνῃ· φθόνος· προσῆ= ↗
 πται· νοσοῦντες· γάρ· ἡμεῖς· οὐχ'· ἡδέως· ό= ↗
 ρῶμεν· τοὺς· ἄλλους· ὑγιαίνοντας· ἀλλὰ· τί= ↗
 ὅφελος· εἰ· πάντες· νοσήσαιεν;· οὐδεὶς· ἀν· ἄλ= ↗

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 Durchschnitt	3,41
GT Training	162
GT Validation	18
GT insgesamt	180



Lat. Tagebuch Martin Crusius, Mh 466-6 (1596-98)

652.	...mp. om̄is principis clonariū p̄o, mifiz. ē t̄um moniti Recensio Gesetzi, M. Evs. Celiq; copioſe 7 ornatē itus, ḡas H̄is. arp Vniuersitati Principi cōmōda. vīz. re. f̄pōta, p̄oſt m̄tia ans. ſorū, p̄oſt adh̄ibitū f̄lī ſorū, p̄oſt adh̄ibitū ſtūng. In alboſ ſuſiōne Gy- poſtāplo. (in vīzia ḡignōn.) illūs pr. A. ſig. p̄oſt Vīzemburg. Dūt Ioan. Frid. 7 alij m̄tia, an einer langen ſaſel. In inferiorre Gyppoſtāplo, dūt m̄tiaſh. In Prio- ri. 7 Gonovariō. Gy ſodim.
647.	g. planer. D. Häbige. D. meyig. D. Harbers D. H̄p̄oſt. a. 10. foreūla. D. Halbitt. b. 8. foreūla. D. ſig. ſig. c. 8. foreūla. M. Crisig. d. 8. foreūla. M. Crisig.

	Crusius_Latin_6a_t
Mh 466-1, S. 8 (26)	3,42
Mh 466-2, 168 (17)	3,76%
Mh 466-3, S. 575 (2,97%
Mh 466-4, S. 110 (3,16%
Mh 466-4, S. 420 (2,96%
Mh 466-5, S. 583 (6,98%
Mh 466-6, S. 83 (7	4,84%
Mh 466-6, S. 635 (6,11%
Mh 466-7, S. 250 (5,10%
Mh 466-7, S. 585 (5,69%
Mh 466-8, S. 26 (1	5,68%
Mh 466-8, S. 512 (4,60%
Mh 466-9, S. 29 (7	5,52%
Mh 198-1, S. 109 (8,25%

CER insgesamt	4,97
Trainingsset	122
Validationset	14
GT insgesamt	136



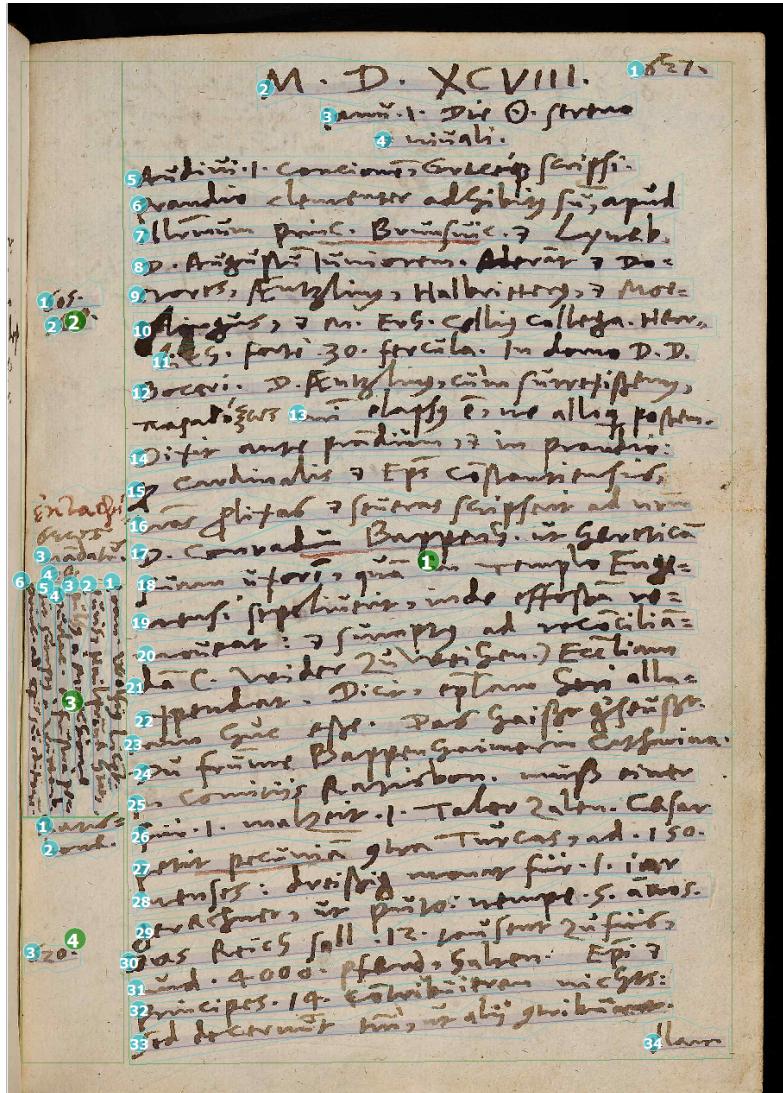
Problematik Transkriptionsrichtlinien

- 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	gqueſt⁹	9
	postque	poſtq̄z	q̄z

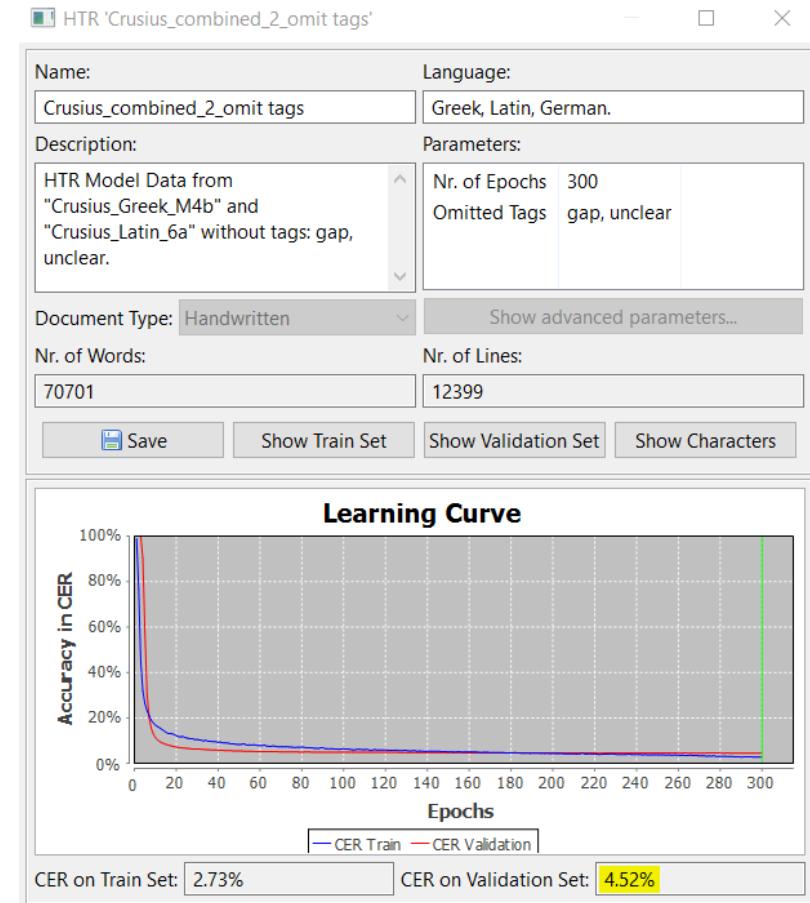


- 1-1 627. ↗
- 1-2 M. D. XC VIII. ↗
- 1-3 Ianu. 1. Die. O. ſereno ↗
- 1-4 7. niūali. ↗
- 1-5 Audiuī. 1. concionē. Græceq; ſcripsi. ↗
- 1-6 Prandio clementer adhibit⁹. fū. apud ↗
- 1-7 illīmum Princ. Brunfuiç. 7 Lynæb. ↗
- 1-8 D. Auguſtū Iuniorem. Aderat. 7 Do= ↗
- 1-9 ctores. Æntzlin⁹. Halbritter⁹. 7 Moe= ↗
- 1-10 glingus. 7 M. Erh. Celli⁹. collega. Herr= ↗
- 1-11 Læh. Forte. 30. fercula. In domo D. D. ↗
- 1-12 Boceri. D. Æntzim⁹. cūm surrexiſem⁹. ↗
- 1-13 m̄. elap⁹. e. ne alloq. poſem. ↗
- 1-14 Dixit ante prādiam. 7 in prandio: ↗
- 1-15 q. Cardinalis. 7 Eps. cōſtantienſis: ↗
- 1-16 Irās plixas. 7 ſeueras. ſcriperit ad. n̄m ↗
- 1-17 D. conradui. Bappenij. ut. hæreticā ↗
- 1-18 ſuum. uxorē. quā. in. Templo. Enge= ↗
- 1-19 nenſi. ſepeluerit. inde. effoſa. re= ↗
- 1-20 moueat. 7 ſinupt⁹. ad. recōciliā= ↗
- 1-21 dā. (wider. zuweihen.) Eccliam ↗
- 1-22 expendat.. Dicit. eptlam. hero. alla= ↗
- 1-23 lam. huc. eſſe. Das. haiſt. gfeuſt. ↗
- 1-24 Du. frume. Bappenhaimerin. Cathaina. ↗
- 1-25 In. comitij. Ratisbon. muß. einer ↗
- 1-26 fur. 1. malzeit. 1. Taler. zalen. Cæſar ↗
- 1-27 petit. pecunia. qtra. Turcas. ad. 150. ↗



Kombinationsmodell Crusius

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

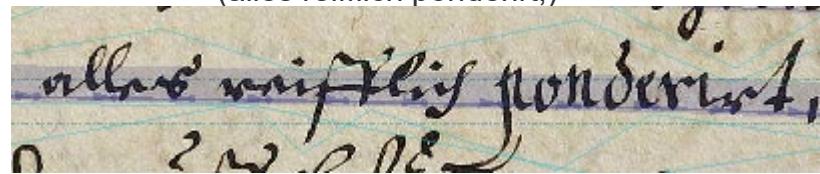




Juristische Konsilien

- 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 reiflich ponderirt,)





Herangehensweise: Erstellung der GT

- Transkription auf Grundlage eines Public Models für Kurrentschrift

HTR Model Data						
All	All engines					
1-9 / 9						
Name	Size	Curator	HTR...	nrOfWor...	Date	
> Transkribus German Kurrent M2	32524	guenter.hackl@...	3209689		27.04.21	
> StAZH_RRB_German_Kurrent_XIX	30919	tobias.hodel@...	26026908		14.03.21	
> Transkribus German Kurrent M2	29820	guenter	3209690		28.01.21	
> Land registers (Verfachbücher) Tyrol, 1750-1800	28866	p.andorfer@g...	129152		21.12.20	
> German_Kurrent_17th-18th	27457	alverman@uni-...	1839841		10.11.20	
> VB_1750_1800_M1Base	26269	p.andorfer@g...	106794		24.09.20	
> German_Kurrent_XIX_pylaia	26068	tobias.hodel@...	5100439		11.09.20	
> German_Kurrent_XIX_M2	23847	tobias.hodel@...	5100439		15.05.20	
> German_Kurrent_XVI-XVIII_M1	19584	tobias.hodel@...	1579208		16.12.19	
> Acta 17 (extended)	32677	alverman@uni-...	1632534		04.05.21	
> Acta_17 PyLaia	27337	alverman@uni-...	594628		04.11.20	
> Acta_17 HTR+	26284	alverman@uni-...	594628		25.09.20	



Modellentwicklung: 1. Modell (26 S. GT)

HTR 'Juristische_Konsilien_Tuebingen_M1a'

Name:	Language:
Juristische_Konsilien_Tuebingen_M1a	German, Latin
Description:	Parameters:
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)	Nr. of Epochs 200
Document Type: Handwritten	Show advanced parameters...
Nr. of Words: 4029	Nr. of Lines: 610
<input type="button" value="Save"/> <input type="button" value="Show Train Set"/> <input type="button" value="Show Validation Set"/> <input type="button" value="Show Characters"/>	

Learning Curve

Accuracy in CER

Epochs

CER Train — CER Validation

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

HTR 'Juristische_Konsilien_Tuebingen_M1a_basemodel'

Name:	Language:
Juristische_Konsilien_Tuebingen_M1a_baser	German, Latin
Description:	Parameters:
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)	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
<input type="button" value="Save"/> <input type="button" value="Show Train Set"/> <input type="button" value="Show Validation Set"/> <input type="button" value="Show Characters"/>	

Learning Curve

Accuracy in CER

Epochs

CER Train — CER Validation

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



Modellentwicklung: Endergebnis (223 S. GT)

HTR 'Juristische_Konsilien_Tuebingen_Eigenauswahl_Testset'

Name:	Language:
Juristische_Konsilien_Tuebingen_Eigenausw...	German, Latin
Description:	Parameters:
gesamte GT; individuelle Auswahl des Validationsets, die alle Einzelgutachten	Nr. of Epochs 250
Document Type: Handwritten	Show advanced parameters...
Nr. of Words:	Nr. of Lines:
41573	6540
<input type="button" value="Save"/> Show Train Set Show Validation Set Show Characters	

Learning Curve

Accuracy in CER

Epochs

CER Train — CER Validation

CER on Train Set: 1.02% CER on Validation Set: 2.48%

HTR 'Juristische_Konsilien_Tuebingen_Eigenauswahl_Testset...

Name:	Language:
Juristische_Konsilien_Tuebingen_Eigenausw...	German, Latin
Description:	Parameters:
gesamte GT; individuelle Auswahl des Validationsets, die alle Einzelgutachten berücksichtigt: UAT 84/13; 25. 35. 73. 111. 192. 289.	Nr. of Epochs 250 HTR Base Model ID 33743 HTR Base Model Name Acta 17 HTR+ ext
Document Type: Handwritten	Show advanced parameters...
Nr. of Words:	Nr. of Lines:
41573	6540
<input type="button" value="Save"/> Show Train Set Show Validation Set Show Characters	

Learning Curve

Accuracy in CER

Epochs

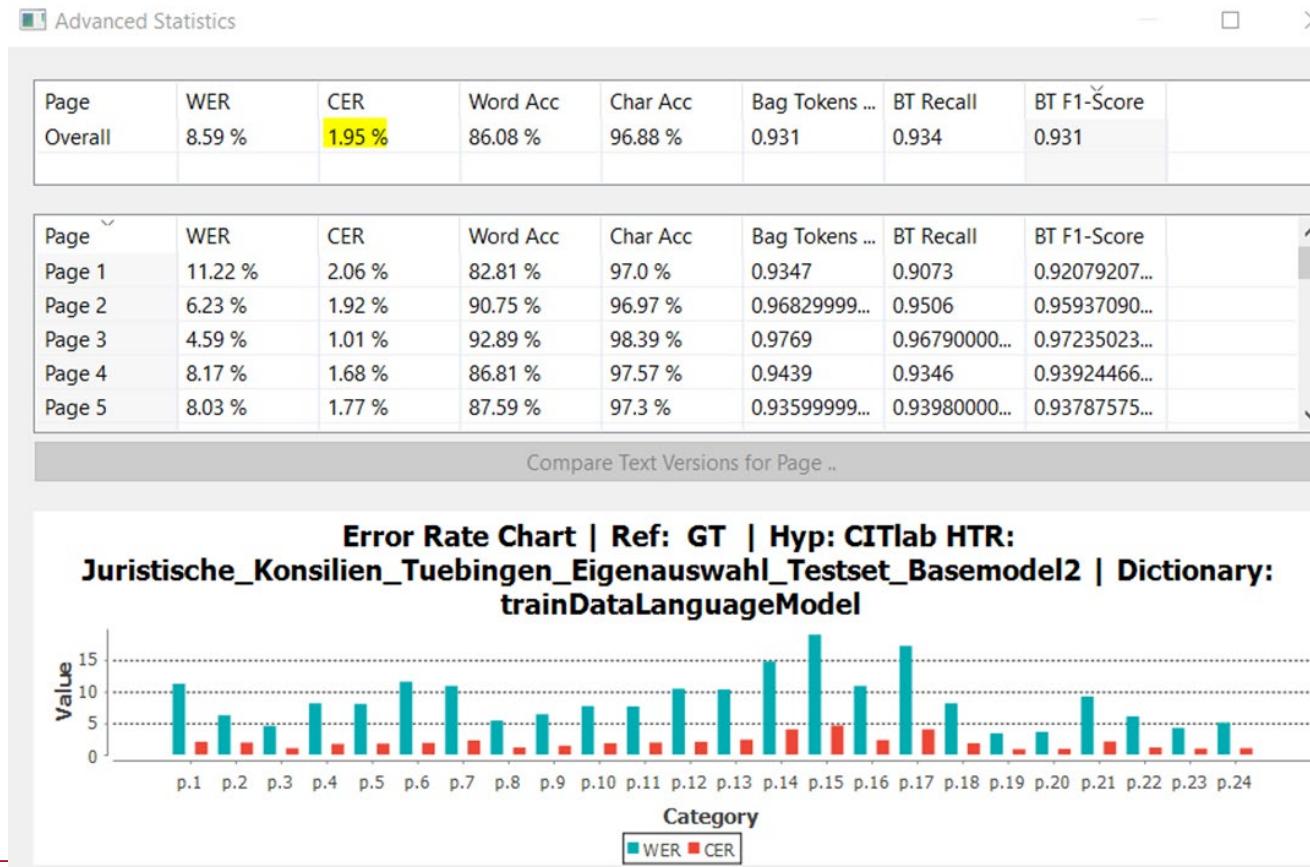
CER Train — CER Validation

CER on Train Set: 0.77% CER on Validation Set: 2.09%



Modell „Juristische_Konsilien_Tuebingen_Eigenauswahl_Testset_Basemodel2“

- durchschnittliche CER mit LM: 1,95%





0,88% CER

*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: adeò, ut, si reus
spondere nolit, & in hac contu
ad veritatem eruendam tortur
Carpz. prax. crim. part. 3. q.
Ubicunq; autem clara & rotunda
ibi non sufficiunt talia verba a
in specie de litis contestatione, q*

459.

Cum tamen Reus, **maximè maxime** in criminalibus, cate,, gorice, **perspicie perspicue** & rotunde ad singula accusatio,, nis membra respondere debeat, ita ut Judex certus fieri poslit, quidnam ab actore probandum sit, & qua,, tenus atq; in quibus membris onus probandi ei injun,, gere debeat; adeò, 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 censemur contestata, si reus respon,, derit: credo **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 responcionem litis contestationem non inducere, docet

Obrecht. tr. de litis contest. cap. 7. n. 6. & 7. ubi subjicit in criminalibus non sufficere credulita,, tis responcionem, 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.

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

4,64%

112.

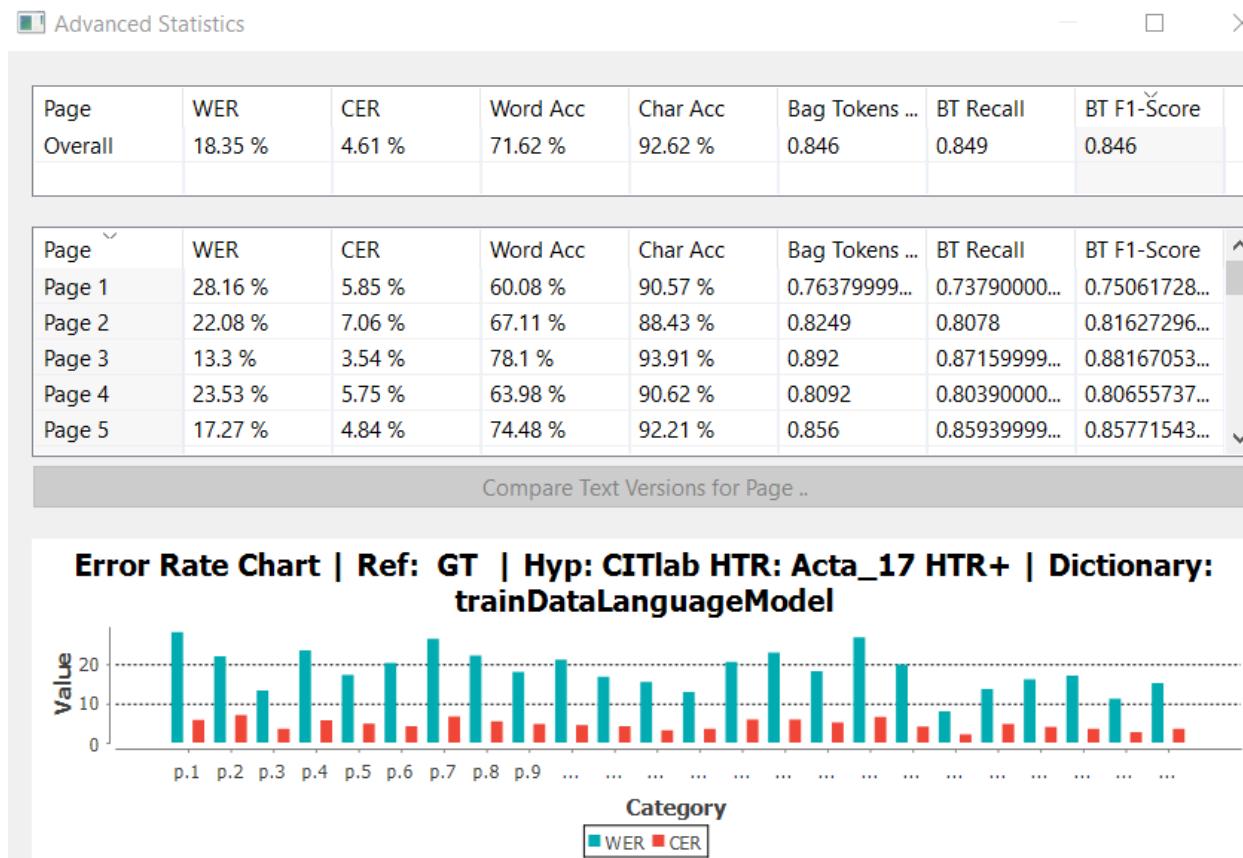
übel tractirt und Gehalten, d3 nicht
ein geringer argwohn sich ereigen
wollen, als ob folches vier Jähriges
Knäblein, Knäblein, durch übele tractation
undt haltung, sein leben Leben eingebuest
hette, und verstorben were, were.

Wie dann hiervon gleichsam fama publica-publica,
und ein allgemein geschrey 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 verstorben Knäblein) habe nit
adlein allein
nicht
igung „gnug zu bsen Eßen bekommen mögen,
sond'n seye auch über die maßen hart
Gehalten, Gehalten, und immer immer zu von des Jacob
Brunen weib gehaft, geschlagen, geschleift,
und im un Christlich tractirt worden. Mit
welchem zeug auf auch
Durehans durchaus
übereinstimmen, Michel
dung Jung der 13. wie auch Melchior Jung
der 14. Item Hanß Georg Schon Schön der 18
Zeug, welche alle, nebenland'n mehr, neben and'n mehr,
einhelliglich, von folcher des Kind
ußgestandener bösen tractation
attestiren thuen. Margaretha
hauß wellen Schneids Tochter Hanß Wolfen Schneid's tochter die
25 gengin zeugin zeigt an, Mann habe d3 Kind
aufgehungert, aufgehängert, biß weiln ihm Brot-brot
eingebrocket, und darüber nur
bloß waßer geschittet, welches d3 Kind
nicht genüßen Können, sondern wid'
von sich geben müssen. dahero vonhard mußen. Dahero Leonhard
fo vorr Ruopp vermeinen undt mittmaßen muttmaßen
thutt, d3 Kindt seye im Leiblein-leiblein undt
mäglin, daureh durch gegebene harte Steiß Speiß
undt tränkh, tranckh, ganzt verderbt worden,



Modell „Acta_17 HTR+“

- durchschnittliche CER mit LM: 4,61%



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.
Ubiq; autem clara & rotund
ibi non sufficiunt talia verba
in specie de litis contestatione,

45 D. 459.

Cum tamen Reus-Reus, maxime in criminalibus, cate
gorice, perspueue-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 subjici possit possit.

Carpz. prax. crim. part. 3. q. 106. n. 22. et 72. & seqq.
Ubiunque 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. I. si debitor-debitor, ff. de heredit. petit.

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

Aecur. Accurs. in I. I. 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. Brunnemann. de process.
cap. 14. n. 15.

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

5,90%

112.

übel tractirt und gehalten, Gehalten, das nicht ein geringen geringer argwohn sich ereigen wollen, als ob solches vier Jähriges venablen, Knäblein, durch übele tractation und haltung, kein leben sein Leben eingebuest 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. Gestaldt same, Gestaldtsame, in specie specie, Michel Kleuth Klenckh sagt: das Kindt puta (puta das verstorbene Knablen) Knäblein habe mit allein nicht iung, gnug zu Eßen bekommen bekommen mögen, sondern wege seye auch über die maßen hert hart Gehalten, und immer zu von des Jacob Brunen weib gehaft, gehast, geschlagen, geschleift, und im un Christlich tractirt worden, worden. Mit welchen Zeugen welchem zeugen auch Durchaus durchaus übereinstimmen, übereinstimmen, Michel Jung der 13. Wie wie auch Melchior Jung den der 14. Item Item Hanß Georg Sehon Schön der 18 Zeuge, Zeug, welche alle, neben anden mehr= andern mehr, einhelliglich, von solcher des Kindt usgestandener usgestandener bösen tractation attestiren thuen. Margaretha hauß wolten Schneids Hanß Wolfen Schneiders tochter die 25 Zeugin zeugin zeigt an, Mann habe das Kin Kind außgehungert, außgehungert, biß weiln ihme Gret brot eingebrohet, eingebrocket, und darüber nur bloß waßer geschittet, welches das Kind nicht genüssen können, genüssen können, sondern wider von sich geben müssen. Dahero vorhand Leonhard Querg Ruopp vermeinen undt mittmaßen muttmaßen thudt, thudt, das Kindt stege seye im leiblein undt mäglin, durch gegebne harte Speiß undt trauern, gantz Verderbt tranckh, ganz verderbt worden,

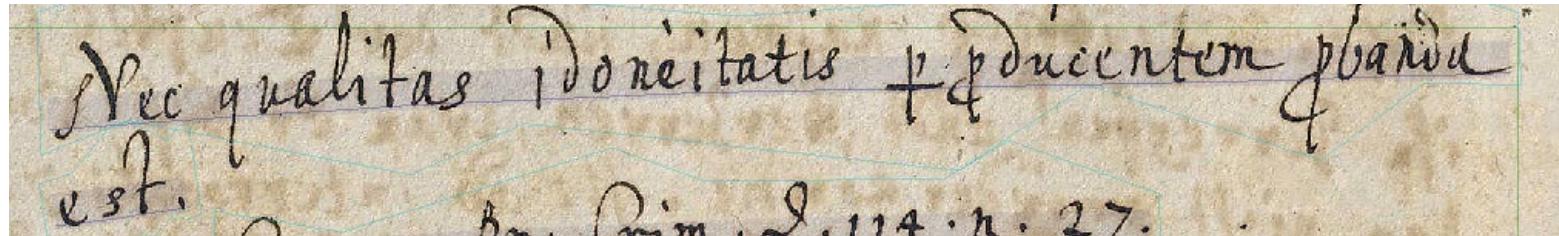
112.

Übel tractirt und gehalten, Gehalten, das nicht ein geringen geringer argwohn sich ereigen wollen, als ob solches vier Jähriges venablen, Knäblein, durch übele tractation und haltung, kein leben sein Leben eingebuest hette, und verstorben were etc. were.

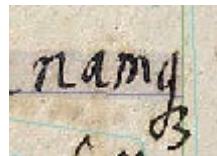
Wie dann hier von hiervon gleichsam fama publica publica, und ein allgemein geschrey entstanden ist, auch die verhörte Zeugen zeugen uno ore bekundtschafften thuen. Gestaldt same, Gestaldtsame, in specie specie, Michel Kleuth Klenckh sagt: das Kindt puta (puta das verstorbene Knablen) Knäblein habe mit allein nicht iung, gnug zu Eßen bekommen bekommen mögen, sondern wege seye auch über die maßen hert hart Gehalten, und immer zu von des Jacob Brunen weib gehaft, gehast, geschlagen, geschleift, und im un Christlich tractirt worden, worden. Mit welchen Zeugen welchem zeugen auch Durchaus durchaus übereinstimmen, übereinstimmen, Michel Jung der 13. Wie wie auch Melchior Jung den der 14. Item Item Hanß Georg Sehon Schön der 18 Zeuge, Zeug, welche alle, neben anden mehr= andern mehr, einhelliglich, von solcher des Kindt usgestandener usgestandener bösen tractation attestiren thuen. Margaretha hauß wolten Schneids Hanß Wolfen Schneiders tochter die 25 Zeugin zeugin zeigt an, Mann habe das Kin Kind außgehungert, außgehungert, biß weiln ihme Gret brot eingebrohet, eingebrocket, und darüber nur bloß waßer geschittet, welches das Kind nicht genüssen können, genüssen können, sondern wider von sich geben müssen. Dahero vorhand Leonhard Querg Ruopp vermeinen undt mittmaßen muttmaßen thudt, thudt, das Kindt stege seye im leiblein undt mäglin, durch gegebne harte Speiß undt trauern, gantz Verderbt tranckh, ganz verderbt worden,



Auswirkung der Transkriptionsrichtlinien



- **Jur_Kons_Tue:** Nec qualitas idoneitatis **p** **pro**ducentem **p**bandu
- **Acta_17:** Nec qualitas idoneitatis **per** **producentem** **p**banda



- **Jur_Kons_Tue:** namq**[ue]**
- **Acta_17:** namque



bloß waßer geschittet, welches d3 Kind
nicht genügen können, sondern wid'

- **Jur_Kons_Tue:** bloß waßer geschittet, welches **d3** Kind
nicht genügen können, sondern **wid'**
- **Acta_17:** bloß waßer geschittet, welches **das** Kind
nicht genügen können, sondern **wider**

analog zum ~~und~~ übereinstimmen, Michel
Dings das 13. Dings aus Melior Dings

- **Jur_Kons_Tue:** welchem zeug **[redacted]** auf Durchanß **übereinstimēn**,
Michel
- **Acta_17:** welchen Zeugen **[redacted]** auch Durchaus **übereinstimēn**, 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