

OCR4all

Erfahrungen und Evaluationsergebnisse der OCR historischer Drucke des 15. bis 19. Jahrhunderts in kritischer Betrachtung

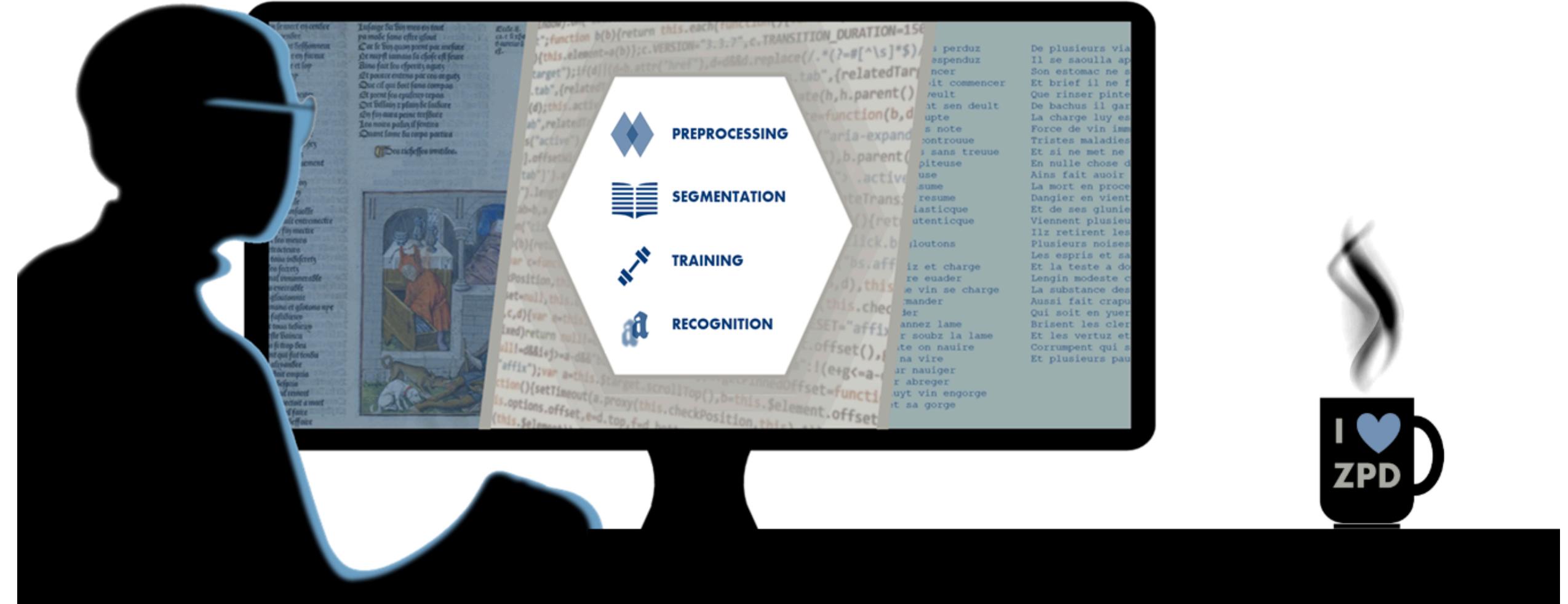
Maximilian Wehner

Zentrum für Philologie und Digitalität „Kallimachos“ (ZPD)
Lehrstuhl für Künstliche Intelligenz und Angewandte Informatik

Universität Würzburg

08.11.2019





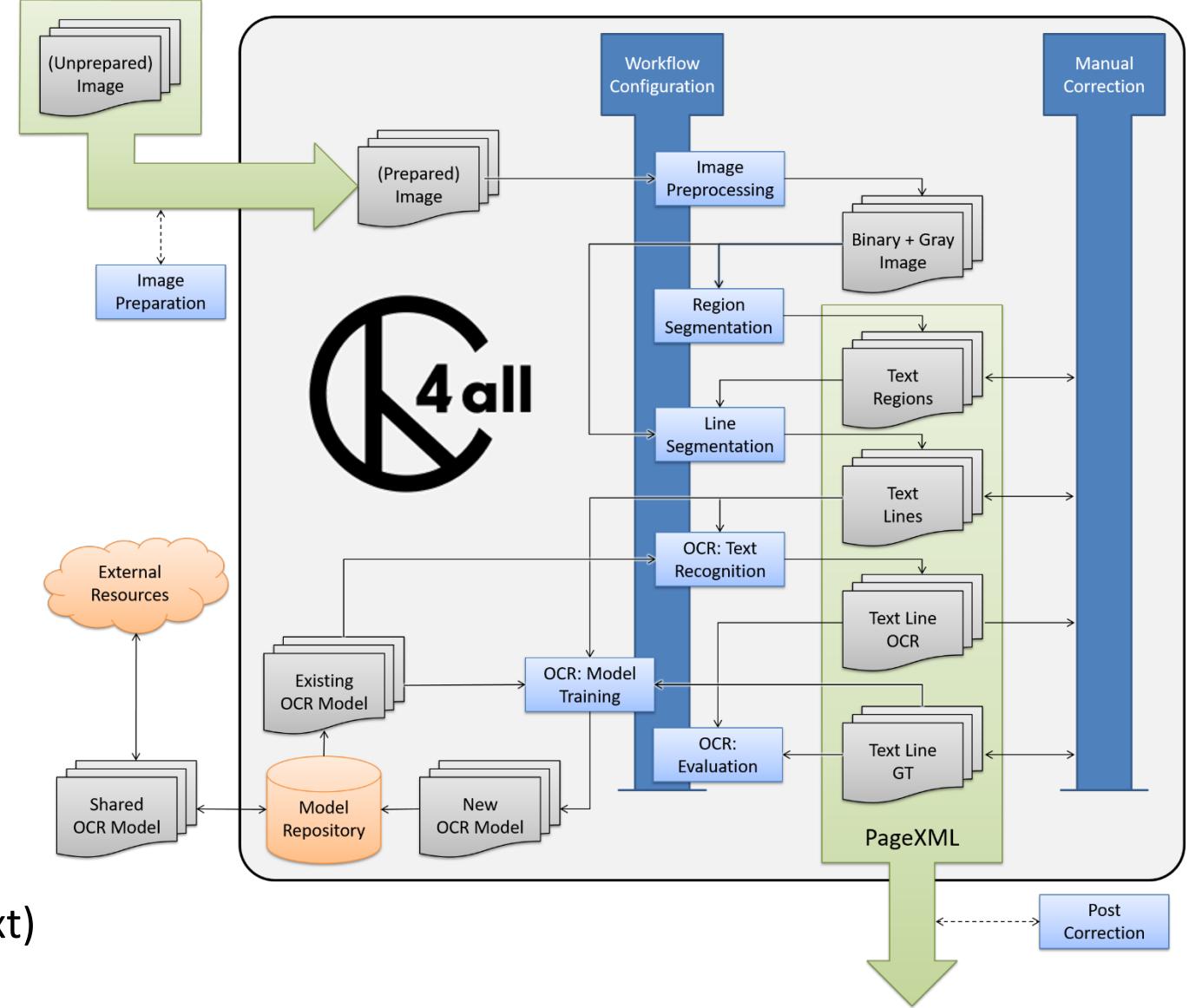
<https://www.uni-wuerzburg.de/en/zpd/ocr4all/>

OCR4all

- Ursprünglich entwickelt für die OCR sehr alter Drucke.
 - Vor einigen Jahren noch als quasi unmöglich erachtet.
 - Sehr komplizierte Layouttypisierung.
 - Aufgrund der höchst varianten Typographie ist ein Training werkspezifischer Modelle unerlässlich.
 - Anspruch: 100% Erkennungsgenauigkeit für Layout und Text.
→ Nutzer nehmen dafür einen gewissen manuellen Korrekturaufwand in Kauf.
- Mittlerweile wird OCR4all für eine ganze Bandbreite unterschiedlicher Text angewendet.
- Weiterentwicklung der Software mit dem Ziel eines höheren Automatisierungsgrades und größerer Robustheit.

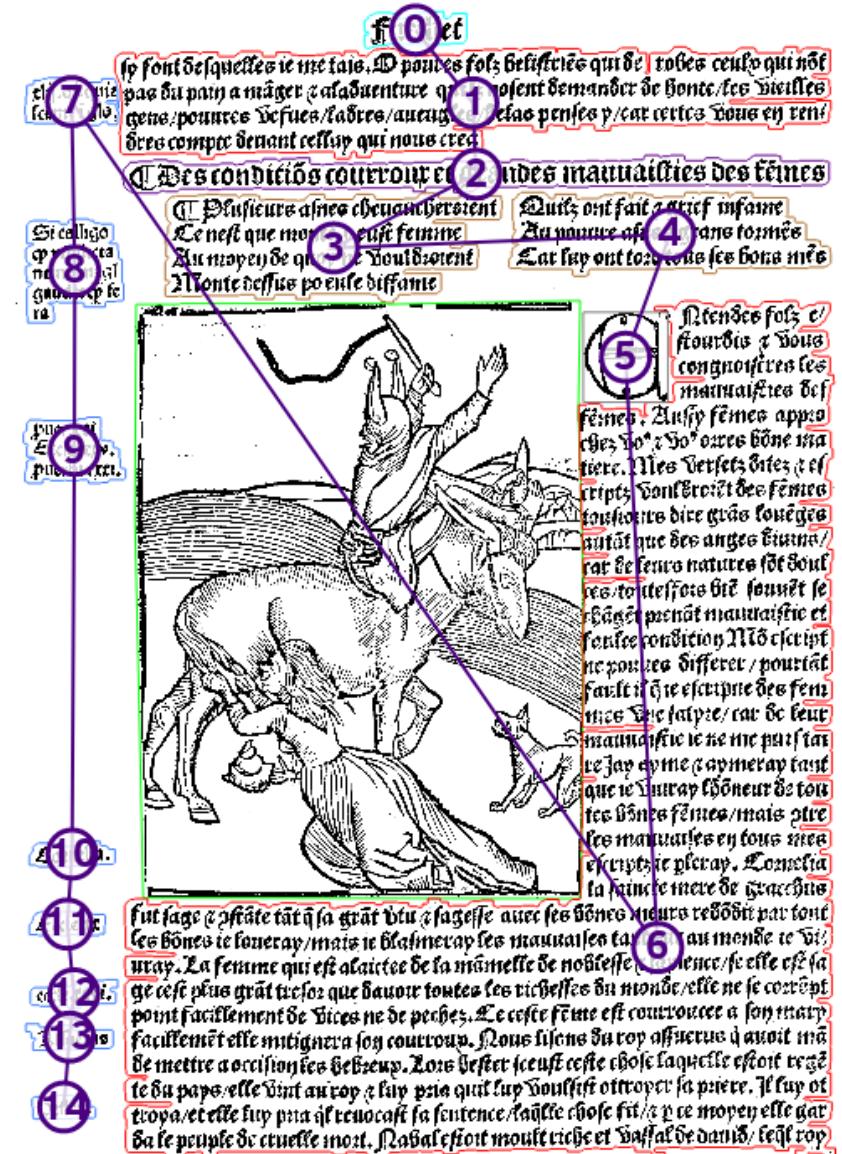
OCR4all – Workflow

- OCropus:
 - Deskewing.
 - Binarisierung.
 - Zeilensegmentierung.
- Calamari:
 - Training.
 - Erkennung.
 - Evaluation.
- LAREX:
 - Regionensegmentierung.
 - Nachkorrektur (Regionen, Zeilen, Text)

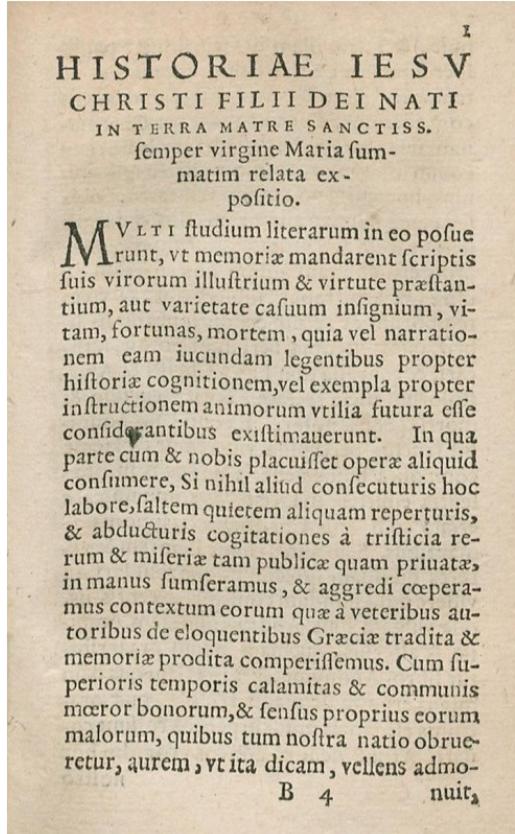


Frühdrucke bis 1600 – Übersicht

- Material
 - 5 Ausgaben des „Narrenschiffs“.
 - 17 Werke des Universalgelehrten Joachim Camerarius.
 - 3 gemischte, früh-neuzeitliche Drucke (Praktikum).
- Anforderungen / Ziele:
 - Zitierfähiger Text (0% Fehler).
 - Präzise semantische Auszeichnung.
- Umsetzung:
 - Bearbeitung in zwei Gruppen:
 - Unerfahrene, nicht-technische Nutzer.
 - Erfahrene Nutzer.
 - Genutzte Instanzen:
 - Lokal bei Einzelnutzer.
 - Serverinstanz an Uni Würzburg.



Frühdrucke bis 1600 – Beispiele I



COMMENTARIORVM LIB. I. 19
BVIT LVMEN LITERARVM, significat deinceps philosophie eloquentiam. Epicuri uero emuli hoc loco notantur, quos & dibi dicit minime malos, sed parum acutos esse. Ian Aristotelis hoc notum est, οὐ γέρων στρατηγός, οὐ δεκατούλαχτης, quem uersum de Tragedia in illam R. Aorem interfecti. Sequitur figura sermoni insitū articula, IN QVAM EXERCITATIONEM OPERAM DEDIMVS. Simile autem est, ut quod dicimus, dare nomen in militiam: uel, quod magis placet, in poteſtatem eſſe, & in honorem, & in discrimen uerſari, de quibus dubitanti iu-
bi DANIELE, aliquando copioſe expofitum ſententiam noſtram. Me-
mini, inquit: & quidem nuper, cum in cūtem sermonem deduxiſſon, pre-
fentibus aliquoſ amicis noſtris, uicindimā fuit mentio quæda habita à te, de
toto genere Ciceroniane imitationis. Sed quis subito tū inde quafi reflitebas,
de inuicta in foro uito dicam neſciō, fuit alioſ permulta quidem requirā.
Omnia de inuicta, inquam, neq; enim ferenda erat et executio harum reu-
illo tempore: & uitabam ijs de rebus ijsq; preſentibus longam diſputationē,
ne, ut nunc huius homines, in malevolentie ſuſpicionem incurrerem. Quin
igitur nunc, inquit, exples defiderium meum, conueniente & in argumento,
& loco libero. Quid nam illud inquam. Cognoscendi elicit, quid ſta-
tus de hac totatione sermonis antiqui, quatenus parum ille cum biderio
iſi congruit, & quatenus huius auctorē, quosq; iſorum potiferm sequen-
dos exigit. Apage inquam: in que difficultates indicis orationem noſtrām, ex hac planitate, quam ingrediſſum est nam uide omnino aliquid uel
la audire nouum. Niſi enim uellem, inquit, habebam quos conſulterem.
Nos uero nibil habemus, inquam, noui. Quare ad illos abeaſtudo. Sed
utros? inquit. Scimus enim diuas contrarie partes conſtituisse: & nouum diſi-
diū eſſe. Hoc uero difficilem, inquam, decernerē eſt. Si tamen omnino de
iudicio meo redi certior cupis, ſcito me planè cum ijs facere, quos conſet
poſt diligentem curam, & ſtudium ardentius, & magis ſpīdum uſum
uerē Latinū demonſtrasse, & alijs directum iter, quo iſi ad normuſſam
Laudem orationis bone & elegantis in hac lingua peruenient. Ex quorum
numero primas tribus P. BE MBO, cuius haec re extat epifola ualde lucu-
lenta. Accepit enim proxime ad ueteres, quorum orationi illius oratio ſimi-
litudo eſt, quo certus de re nota perfeclio, ſibi differere cum putandum.
Sed hec, inquit ille, & legi, & relegam ſepiuſ. Nuic uero te audire cupio.
primū, hoc ſelim edificras, quid sit quod placere tibi tantopere in BEM-
BO diſimilitudinem ueterum. Cupio enim, ut aiebam, te audire de hoc po-

PTRVS. •
Bembus.

6 2 tifſimum

109981/3

Vita ſalutis, homini eſt, is tota mente capiſſat,
Cogiter, edificat, meditetur, corde voluet
Includatq; pio, hac ſe conſoletur, eadem
Pectus ad boſiles armans communiat iſius.
Qv o d patris e gremio celo deſcenderit alto
Filius eterni, ſimul ipſe eternus, & intra
Mortalem celeſte genus conſluerit ortum,
Quen non iſta caput totius machina mundi:
Factus homo in terris, Rerum non indigus ille
Noſtrarum, ſed cura fuit reparare ſalutem
Amiſſam & vitam nobis, ab origine prima
Quos peccatorum duro ſub pondere preſſos
Obruit ira Dei iuſto commota furore,
Hec cauſa in terras celo hunc detraxit, vt effet
Saluator miferorum hominum, Deus uicus &
Rex,
Σωτήρ, λυτερώτης, Σαυτηθός, ἑρα-
βοικτης.

Vita ſalutis, homini eſt, is tota mente capiſſat,
Cogiter, edificat, meditetur, corde voluet
Includatq; pio, hac ſe conſoletur, eadem
Pectus ad boſiles armans communiat iſius.
Qv o d patris e gremio celo deſcenderit alto
Filius eterni, ſimul ipſe eternus, & intra
Mortalem celeſte genus conſluerit ortum,
Quen non iſta caput totius machina mundi:
Factus homo in terris, Rerum non indigus ille
Noſtrarum, ſed cura fuit reparare ſalutem
Amiſſam & vitam nobis, ab origine prima
Quos peccatorum duro ſub pondere preſſos
Obruit ira Dei iuſto commota furore,
Hec cauſa in terras celo hunc detraxit, vt effet
Saluator miferorum hominum, Deus uicus &
Rex,
Σωτήρ, λυτερώτης, Σαυτηθός, ἑρα-
βοικτης.

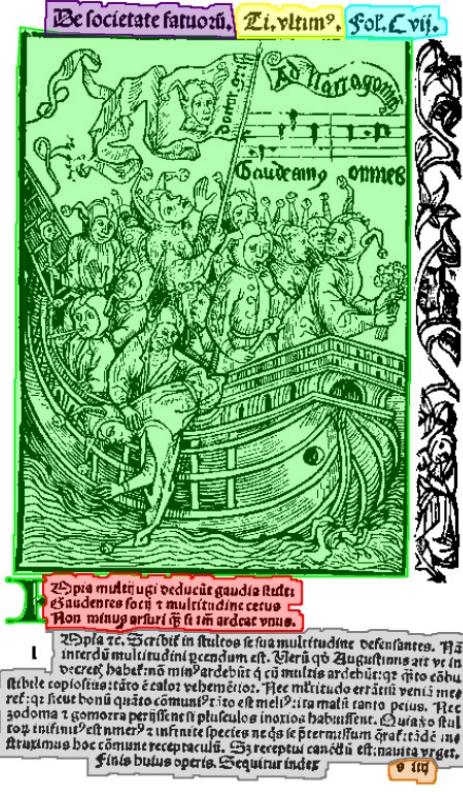
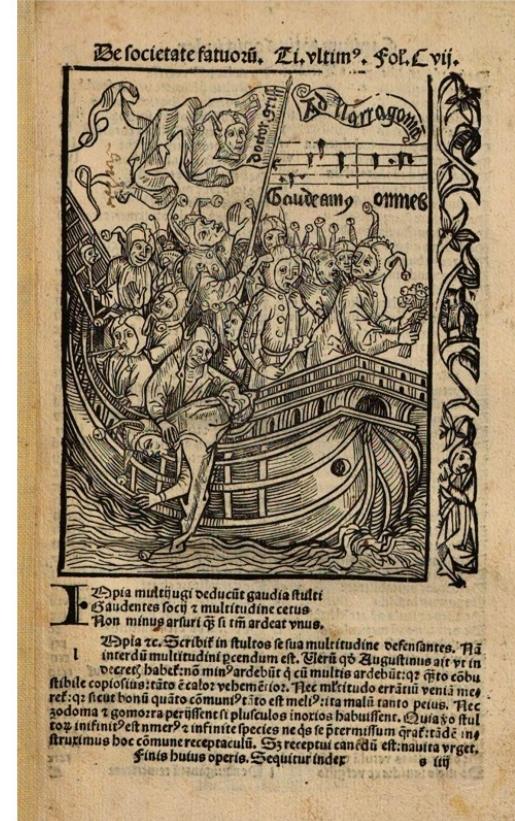
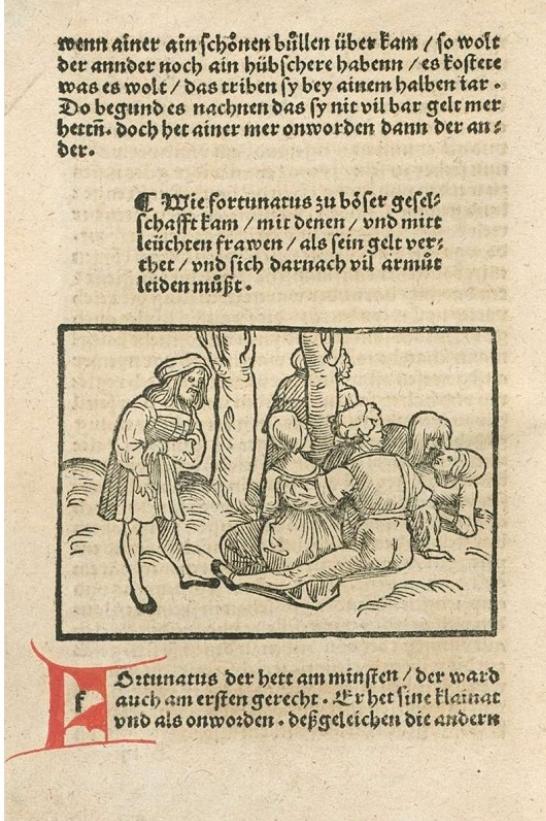
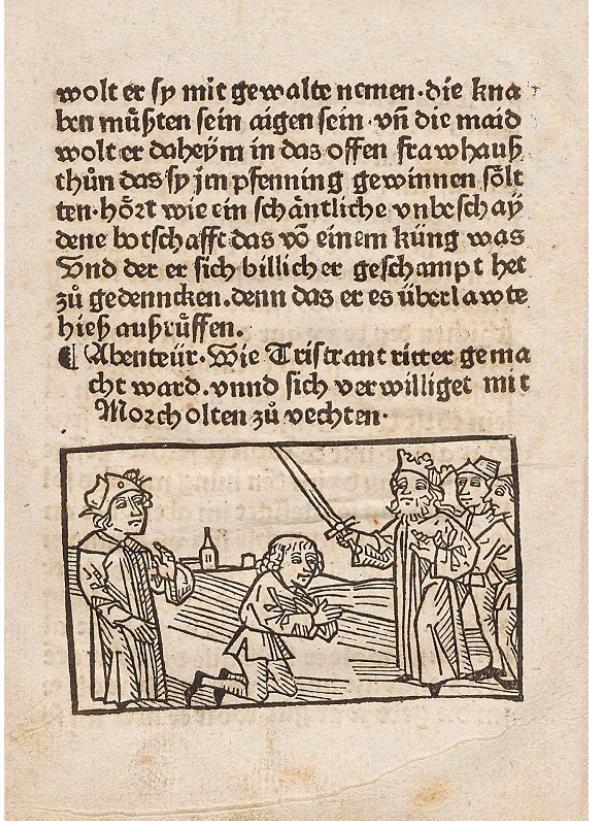
SEQVVNTVR ALII
quidam religiosi argumenti
versus compositi a Io-
achimo Came-
ratio.

K 4 PRE-

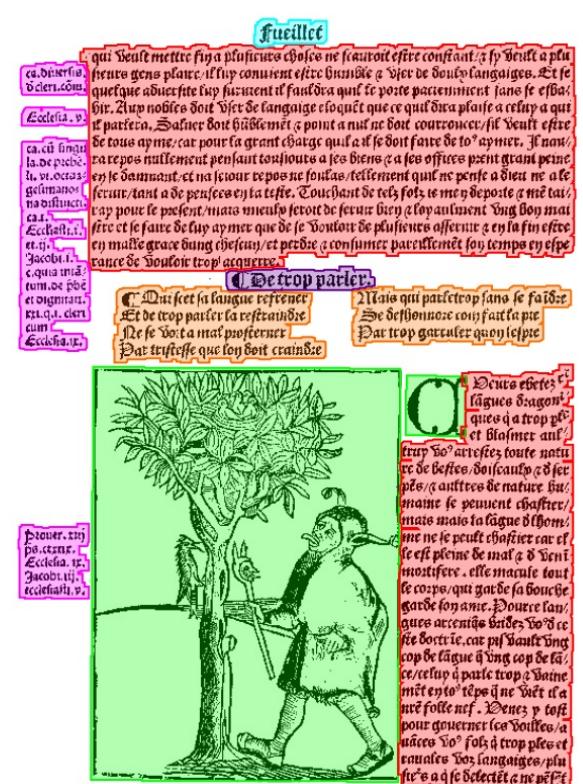
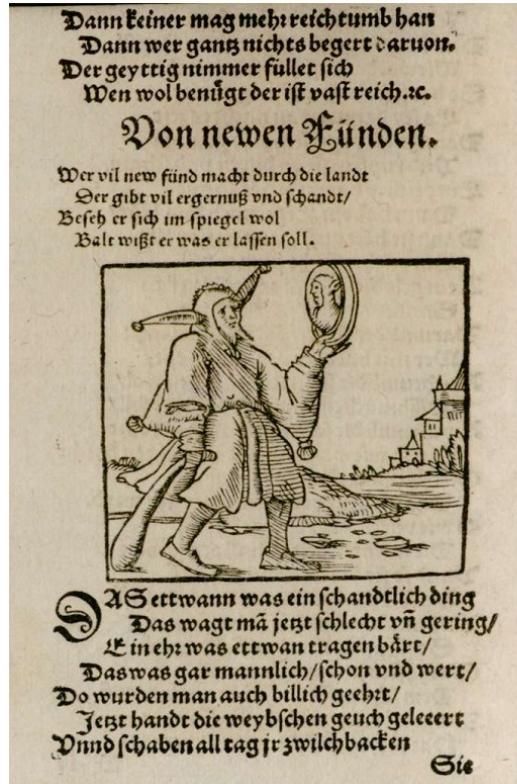
SEQVVNTVR ALII
quidam religiosi argumenti
versus compositi a Io-
achimo Came-
ratio.

K 4 PRE-

Frühdrucke bis 1600 – Beispiele II



Frühdrucke bis 1600 – Beispiele III



Frühdrucke bis 1600 – Ergebnisse

	Unerfahrene Nutzer	Erfahrene Nutzer
Transkribiertes Trainingsmaterial	988 Zeilen	927 Zeilen
Erreichte CER	$0,47\% \pm 0,22\%$	$0,49\% \pm 0,30\%$
Korrekturzeit pro Zeile	$10s \pm 5,2s$	$5,5s \pm 2,4s$
Segmentierungszeit pro Seite	$1,1min \pm 0,5min$	$0,6min \pm 0,2min$

*Jeweils Durchschnittswerte, ggf. mit Standardabweichung

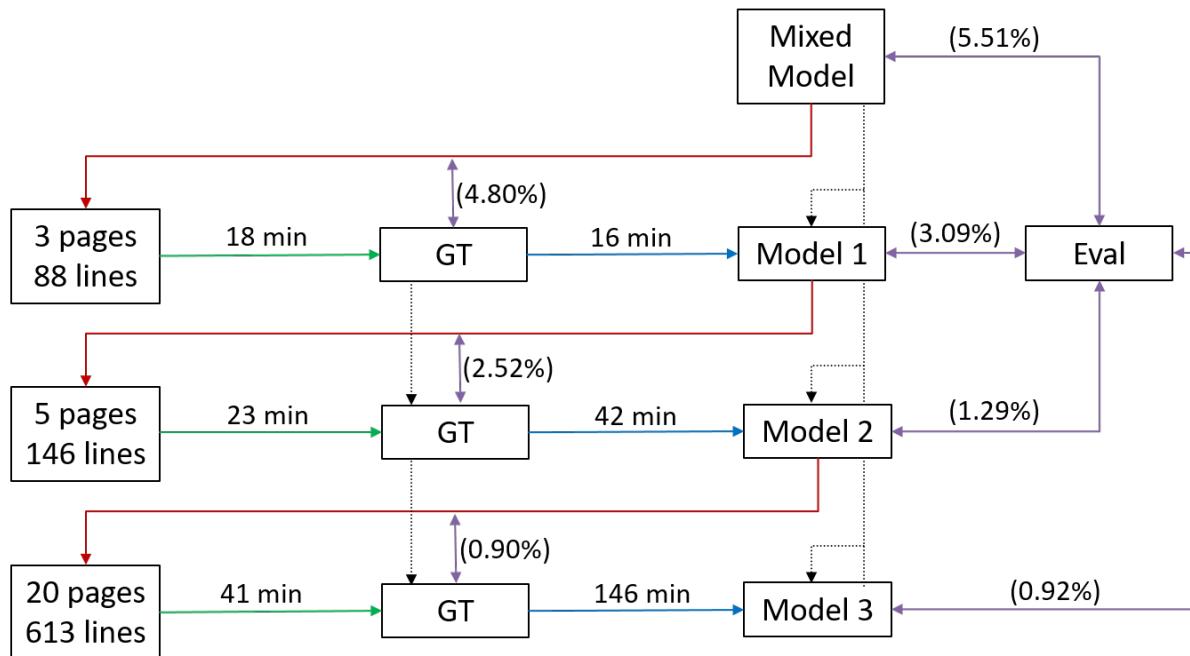
Frühdrucke bis 1600 – Details: Unerfahrene Nutzer

Book Short	User Exp.	Segmentation				OCR				Key Figures	
		#P	#R	t Seg. [min]	t Seg./P [min]	#L	CER [%]	t Corr. [min]	t Corr./L [s]	t All [min]	t All/P [min]
C1532a	1A	55	7	90	1.6	829	0.47	280	20	370	6.7
C1532b	1A	130	7	110	0.8	611	0.73	146	14	256	2.0
C1533	1A	57	5	82	1.4	806	0.20	129	10	211	3.7
C1535	1A	96	7	104	1.1	723	0.39	176	15	280	2.9
C1552	1A	180	6	110	0.6	384	0.20	44	7	154	0.9
C1554	1B	81	6	66	0.8	487	0.36	76	9	142	1.8
C1557	1B	168	5	194	1.2	1,342	0.34	187	8	381	2.3
C1558	1A	94	8	139	1.5	751	0.25	183	15	322	3.4
C1561	1B	344	5	275	0.8	395	0.40	48	7	323	0.9
C1563	1C	158	5	140	0.9	1,175	0.60	95	5	235	1.5
C1566	1D	471	7	370	0.8	596	0.61	48	5	418	0.9
C1568	1B	342	5	223	0.7	406	0.24	36	5	259	0.8
N1494	1E	156	7	210	1.3	2,302	0.69	315	8	525	3.4
N1494	1F	157	7	360	2.3	969	0.82	97	6	457	2.9
N1549	1G	328	7	210	0.6	2,824	0.45	155	3	365	1.1
P1474	1H	198	4	29	0.1	700	0.90	230	20	259	1.3
P1509	1I	218	5	390	1.8	1,501	0.42	310	12	700	3.2
Mean		190	6.1		1.1	988	0.47		10		2.3
StdDev.					0.5		0.22		5.2		1.5

Frühdrucke bis 1600 – Details: Erfahrene Nutzer

Book Short	User Exp.	Segmentation				OCR				Key Figures	
		#P	#R	$t_{\text{Seg.}}$ [min]	$t_{\text{Seg.}/P}$ [min]	#L	CER [%]	$t_{\text{Corr.}}$ [min]	$t_{\text{Corr.}}/L$ [s]	t_{All} [min]	t_{All}/P [min]
C1541	2B	439	8	345	0.8	847	0.92	82	6	427	1.0
C1566	2A	240	7	80	0.3	599	0.57	45	4	125	0.5
C1583	2A	606	7	200	0.3	1,647	1.00	123	5	323	0.5
C1594	2A	420	8	200	0.5	352	0.50	26	4	226	0.5
C1598	2B	344	8	245	0.7	256	0.45	28	7	273	0.8
N1498	2A	161	6	130	0.8	622	0.30	22	2	152	0.9
N1499	2A	166	7	105	0.6	632	0.12	110	10	215	1.3
N1506	2A	215	8	180	0.8	3,161	0.20	-	-	-	-
P1484	2B	372	3	65	0.2	226	0.34	22	6	87	0.2
Mean		329	6.9		0.6	927	0.49		5.5		0.7
StdDev.					0.2		0.30		2.4		0.4

Frühdrucke bis 1600 – ITA



Book Short	User Exp.	It.	New Data					Eval CER [%]	All Data	
			#P	#L	$t_{\text{Corr.}}$ [min]	$t_{\text{Corr.}}$ [s/L]	CER [%]		#P	#L
C1541	2B	1	3	88	18	12	4.80	5.51	3	88
		2	5	146	23	9.5	2.52	3.09	8	234
		3	20	613	41	4.0	0.90	1.29	28	847
		4	-	-	-	-	-	0.92	-	-
P1484	2B	1	5	110	14	7.6	3.53	3.95	5	110
		2	6	116	8	4.1	0.89	1.48	11	226
		3	-	-	-	-	-	0.34	-	-
N1499	2A	1	2	105	65	37	25.22	23.59	2	105
		2	3	138	20	8.7	0.54	2.23	5	243
		3	5	389	25	3.9	1.24	1.63	10	632
		4	-	-	-	-	-	0.20	-	-
C1557	2B	1	4	104	16	9.2	2.00	10.00	4	104
		2	11	307	70	14	6.06	8.64	15	411
		3	15	407	56	8.3	1.60	1.17	30	818
		4	20	524	45	5.2	0.26	0.65	50	1,342
		5	-	-	-	-	-	0.34	-	-
C1558	1A	1	4	125	38	18	15.31	16.86	4	125
		2	8	251	60	14	1.28	0.65	12	376
		3	12	375	85	14	0.58	0.34	24	751
		4	-	-	-	-	-	0.25	-	-
C1566	1D	1	5	122	15	7.4	3.85	4.27	5	122
		2	6	126	18	8.6	3.15	1.45	11	248
		3	12	348	15	2.6	0.22	0.99	23	596
		4	-	-	-	-	-	0.61	-	-

Kräuterbücher des 17. und 18. Jahrhunderts – Übersicht

- Unterstützung einer Lehrveranstaltung für Korpuslinguistik an der HU Berlin.
- Material:
 - Moderates Layout.
 - Teils problematisches Druckbild.
- Anforderungen / Ziele:
 - Vollständige Transkription.
 - Keine semantische Auszeichnung.
- Umsetzung:
 - Dummysegmentierung.
 - Serverinstanz an Uni Würzburg.
 - Studenten korrigieren selbstständig und kollaborativ in Kleingruppen.
 - ITA: Werksspezifisches Training (GPU-Server) zu zwei festgelegten Zeitpunkten.

Kräuterbücher des 17. und 18. Jahrhunderts – Beispiele

Deten WohlEhrwürdigen / Wohl-
EhrenVesten/EhrenVesten/Hoch-Groß-vnd Vor-
Achbaren/Hoch- vnd Wohlgelehrten/Wohl-
weisen/Fürsichtigen/Kunstreichern vnd
Fürnehmern/.

Herrn M. MICHAELI LUDOVICI, der Königl.
Schwed. Mayne, unter vero Höchstölichen Haubt-
Armeē im Leutschland wohlerordneten General-Su-
perintendenten vnd des FeldConſiliorij Präfidenten.

Herrn MICHAELI Siegenſpeck / Jurisperito, Fürſl.
Sächs. wohlbefesteten Reichsreichen Verwaltern / wie
auch des J. S. Gymnaſij h. Scholaren, vnd dann
der J. S. Apotheken Inſpectori primario zu Coburg.

Herrn GEORGIO PHILIPPO Aibling / Juris-
perito, Fürſl. Sächs. wohlbefesteten Amptioſchöfferer /
vnd des J. S. Gymnaſij Scholarchen, wie auch der

J. S. Apotheken Corporis Inſpectori aliiſter.

Herrn NICOLAO Vollmahn / Fürſl. Sächs. wohlbefestigten Amptioſchöfferer aliiſter.

Herrn ANDREÆ Erhard / L. L. S. S. Cand. wohlbefestigten Bürgermeiſtern aliiſter.

Herrn JOHANNI Aufzourm / Fürſl. Sächs. wohlbefestigten Rent- vnd Hammermeiſtern aliiſter.

Herrn JOHANNI Schaden / Fürſl. Sächs. wohlbefestigten Steuermeiſtern aliiſter.

Herrn JOHANNI Weiland / Fürſl. Sächs. wohlbefestigten LandCaliſtern aliiſter.

Herrn JOHANNI Hoffmann / des Rathes vnd Fürſl.
S. wohlbefestigten Hof- und Stadt-Apothekern aliiſter.

Herrn SEBASTIANO May / des Rathes vnd wohlbefestigten Untergeldiern aliiſter.

Mit denen insders Großgäntigen Herrenquaten
Gönner/ Freunden vnd Schwägers zu sonderlich
ihren Ehren und Gefallen offeneret vnd dedicirt
G. B. D. p. f. F. F. & in Part. Polaris.

Staatsbibliothek
München



Dedicatio.

Damit aber ſolche zubereitung recht
geschehen möge / fo hat Gott der H E R O
durch ſein ewigen / den Menschen das
Liecht der Natur geoffenbart / daß durch
die Kunſt Chymiam vnd Spagiricam,
oder ſchende Kunſterfundn worden / wie
man auf den Vegetabilien, Animalien
vnd Mineralien, iſhre kraft zu extrahiren,
die ſubtile Olliteren vnd Balsamen
zu destilliren, die Mineraliſchen Salia
vnd Crocum zu reuerberieren vnd zu
figiren; vnd alſo das reine von dem unrei-
nen zu ſepariren vnd zu ſchenden / ſinte-
mal durch dieſelbe Kunſt die Materia von
ihren Ordibus, vreinigkeit vnd von ih-
rem Gifft gescheiden wird / wegen ſeiner
ſubtilen præparationem.

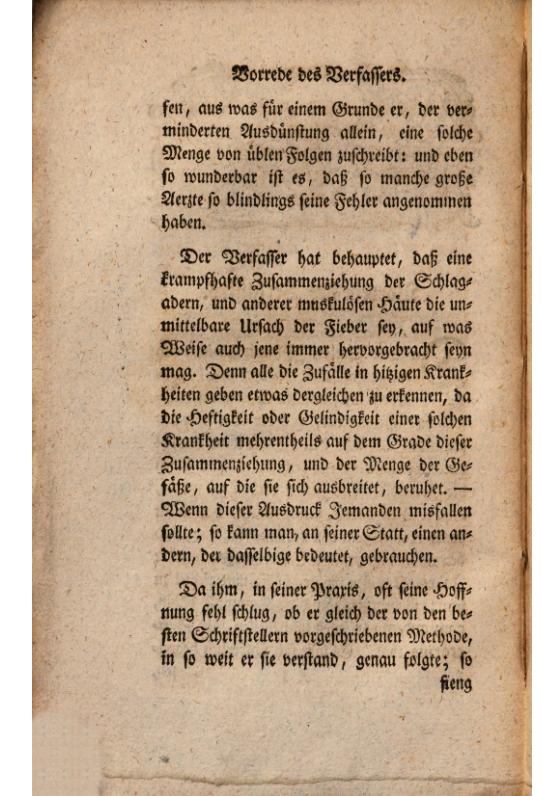
Dah ich aber E. Fürſl. Gn. dī gerin-
ge Tractetlein habe jüngſchrieben / hat
mich erſtlichen darzubewegen / die Kunſt
digewillfahrung vnd Gutthat / ſomit
von E. F. Gn. viel Jahr hero Gnädig wi-
derfahren / auch das E. Fürſl. Gn. je vnd
allezeitſolche Difſillir- vnd ſcheide Kunſt
geliebet / dieſelbige in hohen Ehren vnd
Würden

Vorrede des Verfassers.

fen, aus was für einem Grunde er, der ver-
minderten Ausdünftung allein, eine ſolche
Menge von übeln Folgen zuschreibt: und eben
ſo wunderbar ist es, daß ſo manche groſe
Ärzte ſo blindlings ſeine Fehler angenommen
haben.

Der Verfasser hat behauptet, daß eine
krampfhafte Zusammenziehung der Schlag-
ader, und anderer unruhigen Häm̄e die un-
mittelbare Ursach der Gieber ſey, auf was
Weife auch jene immer hervorgebracht ſey
mag. Denn alle die Zufälle in hiſigen Krank-
heiten geben etwas dergleichen zu erkennen, da
die Heftigkeit oder Gelindigkeit einer ſolchen
Krankheit mehrtheils auf dem Grade dieſer
Zusammenziehung, und der Menge der Ge-
fäße, auf die ſie ſich ausbreitet, beruhet. —
Wenn dieſer Ausdruck ſemanden mißfallen
ſollte; ſo kann man, an ſeiner Statt, einen an-
dern, der dasselbige bedeutet, gebrauchen.

Da ihm, in ſeiner Praxis, oft ſeine Hoff-
nung fehl ſchlug, ob er gleich der von den be-
ſten Schriftstellern vorgeschriebenen Methode,
in ſo weit er ſie verſtand, genau folgte; ſo
ſtieg



Kräuterbücher des 17. und 18. Jahrhunderts – Ergebnisse

Werk	fraktur_hist	19th_cent	WSp1	WSp2
K1625	89,95%	37,96%	95,51%	98,47%
K1650	85,41%	45,32%	99,16%	99,11%
K1700	86,15%	63,71%	95,57%	95,57%
K1718	88,36%	61,83%	90,69%	97,75%
K1743	83,49%	76,70%	95,38%	98,30%
K1773	88,09%	66,51%	96,68%	97,83%

Frakturromane 19. Jahrhundert – Übersicht

■ Material:

- Moderates Layout.
- Relativ standardisierte Typographie.
- Insgesamt besserer Erhaltungszustand.

■ Anforderungen / Ziele:

- „so gut wie möglich“
- Kein werksspezifisches Training.
- Keine semantische Auszeichnung.

■ Umsetzung:

- Vollautomatischer Ansatz.
- Dummysegmentierung.
- Erkennung mit sehr starkem, passenden gemischten Modell
(Calamari-Modellensemble „fraktur_19th_century“).

Frakturromane 19. Jahrhundert – Beispiele

140

und erhob die Schönheit der Blume, und ihre Freude daran mit kindlich warmem Gefühl; aber der Ton ihrer Stimme, die sterbenden Blicke des schönen Auges, die zuweilen mit klappendem Ausdruck auf mich fielen, zeigten deutlich, wie schmerzlich ihr Herz berührt war. Leonore kürzte den Besuch ab, und trachtete unter dem Vorwand vieler Geschäfte nach Hause; mir aber ließ sie völlige Freyheit, zu thun, was ich wollte. Ich fühlte ihre Absicht, aber mein Entschluß war fest. Ich ging, und schrieb an Rosalien, so ruhig und doch so herzlich, als ich es vermochte. Ich setzte ihr die Unumgänglichkeit meiner plötzlichen Abreise aus häuslichen Rücksichten und mit einem gläubischen Vorwande auseinander, Leonorens gedachte ich mit keiner Sylbe.

Am andern Morgen brachen wir früh auf. Rosalien habe ich nicht wieder gesehen.

Hier hat mich gleich eine Menge durch lange Abwesenheit gehäufte Geschäfte empfangen. Ich habe sehr viel zu thun, und bin dessen froh. Im rajhen Wirken und Schaffen fühlt sich die echte Kraft, und vergißt sich mancher thörichte Gedanke. Auch ist es ein belohnendes Gefühl, am Schlusse jedes Tags, jeder Woche

„Ah, Freundin, wenn es so fortfährt, so muß ich eine Bette wählen, um glücklich zu seyn! — Hart, — ha! wie fällt mir der Name ein! — Wie ich zittere! — Seit drei Tagen ist es das erstmal, daß ich ihn ausspreche. Denn er ist schon drei Tage frank, und kommt nicht von seinem Zimmer.

Was doch dem lieben Jünglinge fehlen mag? Vorgestern war er mit mir noch im Garten; weinte mit mir — o, ich vergesse die Zärtlichkeit nie, mit der er weinte, mich ansah, mich gern getröstet hätte; — wir giengen mit einander fort; — und seit dieser Zeit läßt er sich nicht sehen. Mich dauert der gute Hart sehr. Er hat so viel Edles, so viel Gefühl, daß er es verdient, bedauert zu werden.

Es steht jemand! — Vermuthlich wieder eine Visite. —

Abends um 10 Uhr.

„Ah, Julchen! Hart war bes mit. Ich weiß nicht, warum ich so erschrat, als ich ihn sah. Er erkundigte sich um mein Wohlsein, bedauerte mich, daß ich unpaß wäre, (denn ich mußte schon — Ehrenhalber Kopfweh vorshützen) und lachte durch tausend Kleinigkeiten, mich zu zerstreuen. Aber ich las es in seinem ganzen Vertragen, daß er einen gewissen Grad von Zurückhaltung und Zwang angenommen hatte. Wor-

R 5

153

— 6 —

Das Mädchen ging singend über den Vorsaal, und ihre schönen Töne zogen alle in sein Herz. Er verliebte sich in die schlanke, bewegliche Gestalt, in den kleinen Fuß, in den runden weißen Arm, strich vor Strich bis in das goldene Haar hinauf, das über dem schönsten Oval des Gesichts sich lockte.

Das Mädchen sah ihm nach, wenn er sich auf sein Pferd schwang; die edle, hohe Gestalt, die schwarzen, blühenden Augen, und gar, wenn er nun seine Hand an den Hut hielt, sie zu grüßen — kurz nach sechs Wochen war das erste Gespräch auf dem Vorsaal, nach acht Tagen der erste Seufzer mit seinem Echo, nach acht Tagen der erste Händedruck, der dem Kornet vorfiel; als würde er von Neuem geadelt; nach einem halben Jahre erst, was sagt Ihr zu dem Mädchen? da der Kornet sie aus der Komödie führte, der erste Kuß auf dem dunklen Vorsaal.

Kurz der Kornet hätte das Mädchen auf der Stelle geheirathet; so flamme dieser Kuß in seiner Seele. Er dachte

— 36 —

— 37 —

Glück auf diese Seite zog. Er sah ein junges Mädchen, in Trauer gekleidet, in der furchtsamsten und demuthigsten Stellung da stehen; unter der schwarzen Florhaube wallten goldne Locken über die lichte Stirn. Es war ein unbeschreiblich liebliches Gesicht, eine höchste, zarte, schlanke jungfräuliche Gestalt, der die demuthige Furchtsamkeit neue Reize gab.

„Warum nicht früher?“ rief mit harter Stimme Genf.

„Ich stehe schon seit einer Viertelstunde hier!“ antwortete leise die schone Stimme.

Genf winkte weg, und das Mädchen verschwand eben so ungehört, als sie gekommen war.

„Sehen Sie, Herr Nachbar,“ sagte Eugenie lächelnd — „auch Stumm belebt Ihre Erzählung zu Seufzern.“

„Stumm?“ sagte er, und schwieg.

Das Gespräch hob wieder an; aber der Kapitän war ein wenig zerstreut.

sehr unglücklich. „Als ich ihm nahe genug war, und ihn anzureden, und den zusammengelegten Brief wieder dazurücken, sagte ihm meine Erzählerin: Theurer, gütiger Onkel! was hast du George gehabt?“ fragte ich aufsäugend.

Was zu deiner und meiner Nähe gehörig gewesen wäre, ist ein besondres von mir. Wie sehr über Onkel Knasch mich zu schämen ist!

Karl ist mir nichts entgangen von alldem was deine Seele beschäftigte. Ich sah, wie aufrichtig du dich bemütest, deine unangenehmen Gefühlsungen zu überwinden.

Ich wolle sehen, wie weit dein Willen und deine Kraft gehen würden. Der erste war stärker als die letzten; und dann wünschte ich auch zu wissen, ob die gefährliche Verwohnerin der Hoffnungslaus auch an dich denkt und deine Unabhängigkeit verdient.

Frakturromane 19. Jahrhundert – Ergebnisse

Book	ABBYY	OCR4all	ErrRed.
F1781	2.9	0.60	79.3
F1803	27	4.89	81.9
F1810	3.8	0.61	84.0
F1818	10	1.35	86.6
F1826	1.1	0.06	94.4
F1848	0.93	0.20	78.5
F1851	1.0	0.16	84.0
F1855	4.0	0.33	91.8
F1865	1.6	0.18	88.8
F1870	0.48	0.13	72.9
Avg.	5.3	0.85	84.2

ABBYY				OCR4all			
GT	OCR	CNT	%	GT	OCR	CNT	%
„		64	2.6	„		63	11.9
s	S	57	2.3	n	u	14	2.7
,		57	2.3	f	s	12	2.3
e	c	50	2.0	i	l	12	2.3
e	"	40	1.6	r	t	12	2.3
s	r	40	1.6	"	,	9	1.7
-	"	40	1.6	,		9	1.7
.		39	1.6	i		8	1.5
x	"	36	1.5	c	e	8	1.5
		35	1.4	,		6	1.1
Remaining				Remaining			
81.6				71.1			

Zusammenfassung und Diskussion

- OCR4all vielseitig einsetzbar.
 - Erfolgreiche Verarbeitung historischer Drucke des 15. bis 19. Jahrhunderts.
 - Hauptanwendung: lokale Installation beim Nutzer.
 - Nutzung als Serveranwendung prinzipiell bereits jetzt möglich.
- Ergebnisse und Aufwand stark abhängig vom
 - Material.
 - Automatische Segmentierung möglich?
 - Passendes gemischtes Modell vorhanden?
 - ...
 - Nutzeranforderungen.
 - Semantische Auszeichnung nötig?
 - Reading Order trivial?
 - Ansprüche hinsichtlich Genauigkeit?

(Aktuelle) Weiterentwicklungen

- Abrunden der aktuellen Version.
- Virtual Box als zusätzlicher Weg der Einzeldistribution.
- Ausbau zu einer vollwertigen Serverapplikation.
 - Nutzerverwaltung inklusive der Möglichkeit mehrstufiger Rollenzuweisung.
 - Umfassendes Ressourcenmanagement inklusive voller Unterstützung der bestehenden GPU Infrastruktur.
 - Ggf. Ausbau zu ZPD-Service.
- Stärkere Nutzung von und Orientierung an OCR-D Lösungen.
- Konfidenzbasierte, interaktive Nachkorrektur.
- Leistungsstärkere vollautomatische Segmentierung.