

Transkribus

Eine Forschungsplattform für die
automatisierte Digitalisierung, Erkennung
und Suche in historischen Dokumenten

Günter Mühlberger

Universität Innsbruck,

Digitisation and Digital Preservation Group

The logo consists of the word "READ" in a bold, dark blue, sans-serif font. The letters are thick and closely spaced, with a slight shadow effect on the right side of each letter.

Agenda

- Einleitung
- Technologie
- Resultate
- Transkribus – Expert client
- Transkribus – Plattform
- Sharing is caring oder die Zukunft von Transkribus

Einleitung

File Edit View Chronology Markers Extras Help

Flüge nach... Neuer Tab Fahrplana... Flüge nach... TextImage... Text2voro... textImager UW Philtag 14... http://u...hei.nz/ ToggI - Ti... READ READ ... X READ_D7.7_HT...

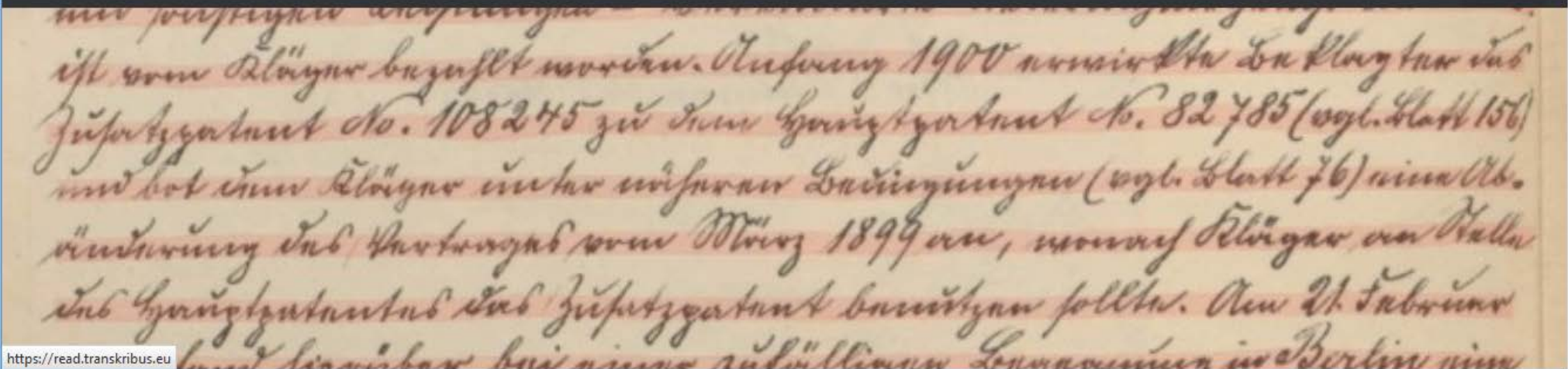
https://read.transkribus.eu 120% textimager hu

Google Kalender SP DEA ACM ulb PRIMO tS WIKI EU_NEWS Solritas TSX Adobe Connect Login PARTHENOS PartPortal Transkribus Wiki READ Wiki myPayLife - Übersicht SBB aktuell

READ

Recognition and Enrichment of Archival Documents

Home News Mission Network Research Services Publications Contact



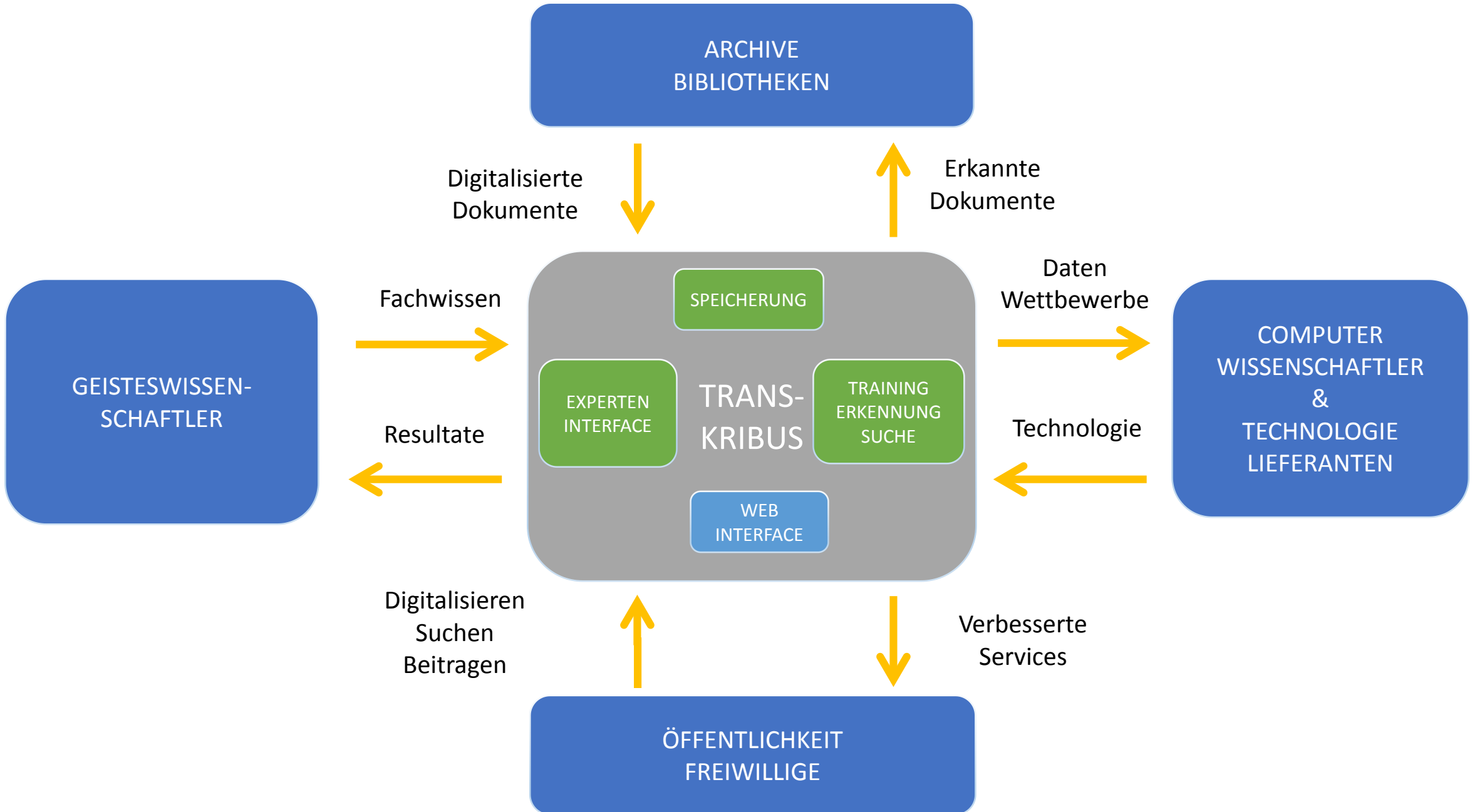
ist vom Klöner bezuflit worden. Auf dem 1900 nummerirten Klöner ist
Zufatzgutacht No. 108245 zu dem Hauptgutacht No. 82785 (vgl. Blatt 15b)
mit dem Klöner in der wifurten Sachverhänge (vgl. Blatt 7b) eine Ab-
weichung des Hauptgutacht vom März 1899 an, wovon Klöner von Halle
das Hauptgutacht des Zufatzgutacht bewähren sollte. Am 21. Februar
wurde hierüber bei einer öffentlichen Lesung in Berlin eine

https://read.transkribus.eu

Desktop c60817 Links DEU 06:59

READ - Transkribus

- Fakten und Zahlen
 - Horizon 2020 Projekt der Europäischen Kommission
 - 8,2 mill. EUR Förderung
 - Projektdauer: 1.1.2016 – 30.6.2019
 - 14 Partner, koordiniert von der Universität Innsbruck
- Schwerpunkte
 - Research: 60% - Pattern Recognition, Machine Learning, Computer Vision,...
 - Netzwerkbildung: 20% - Wissenschaftliche Wettbewerbe, Workshops, Support,...
 - Service: 20% - Aufbau einer Serviceplattform
- Serviceplattform: Transkribus
 - Digitalisierung, Transkription, Erkennung von und Suche in historischen Dokumenten ermöglichen
 - Forschungsinfrastruktur



Technologie

Texterkennung

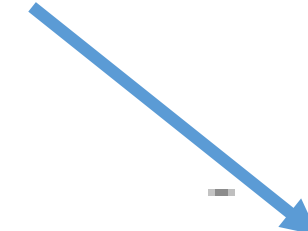
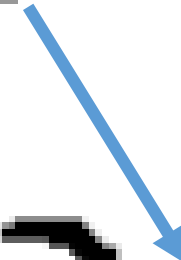
Be r a n f a l t u n g

Be

r

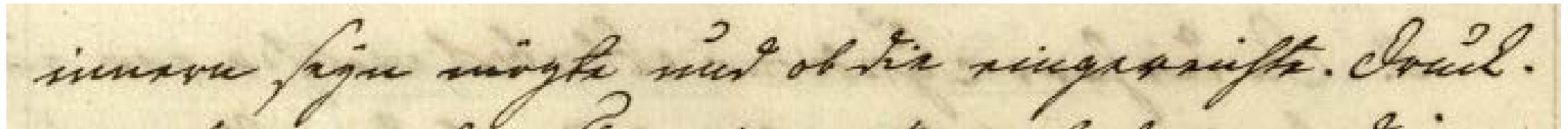
a

n



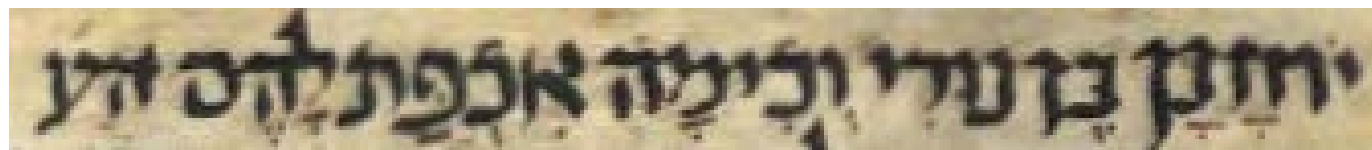
und kluge Veranstaltung/ des Käyserl. General Feld-Marschall Lieutenants

und kluge Veranstaltung/des Käyserl. General Feld=Marschall Lieutnants



innere seyn mögte und ob die eingereichte. Druck.

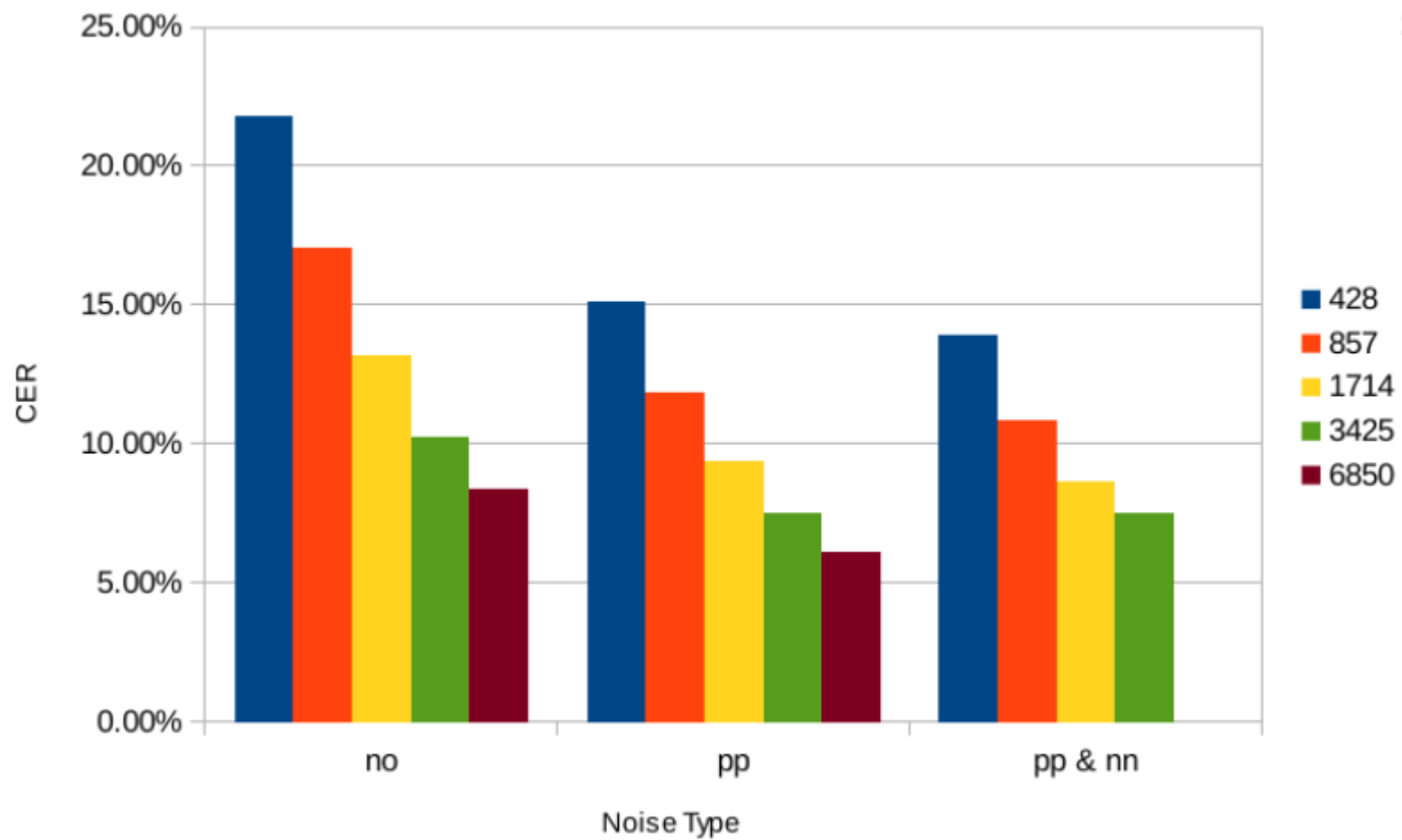
innere seyn mögte und ob die eingereichte. Druck.



יִחוּס בֶּן נוֹרִי וְכִי מָה אֲכַפֵּת לָהֶם הֵע

יִחוּס בֶּן נוֹרִי וְכִי מָה אֲכַפֵּת לָהֶם הֵע

Influence of Number of Lines on CER



Source: Gundram Leifert (CITlab)

Fortschritte im READ Projekt – seit 2016

Dataset	SPRNN (=2016)	HTR+ (e2017)	HTR+(e2018)
StAZH	14,48*		
Bozen Ratsprotokolle	(24,39)		

- All figures as CER – Character Error Rate
- No dictionaries

Source: CITLab team

Fortschritt 2017 (nicht implementiert)

Dataset	SPRNN 2016	HTR+ (e2017)	HTR+(e2018)
StAZH	14,48*	4,45	
Bozen Ratsprotokolle	(24,39)	6,70	

- All figures as CER – Character Error Rate
- Source: CITLab team

Fortschritt Ende 2018 (implementiert)

Dataset	SPRNN 2016	HTR+ (e2017)	HTR+(e2018)
StAZH 19 th C.	14,48*	4,45	2,97
Bozen 17 th C.	(24,39)	6,70	4,89

- All figures as CER – Character Error Rate
- Source: CITLab team

Zeilenerkennung

Zeilenerkennung

- Zu Beginn des Projekts “die” Herausforderung
 - Komplexes Layout
 - Viele verschiedene Texttypen
 - Tabellen
 - Schmutz, Ränder, Durchscheinen, Streichungen, Zeichnungen, Separatoren, Buchschmuck,...
- Erfolgsgeschichte
 - Neue Methoden: Machine Learning
 - Representatives und gut ausgewähltes Datensatz (2000 Seiten von verschiedenen Archiven und Zeitepochen)
 - Offshore Firma zur Erstellung des “Ground Truth” (100.000+ Zeilen manuell eingezeichnet)
 - Wissenschaftlicher Wettbewerb: ICDAR 2017
 - Forschungsdaten sind auf ZENODO verfügbar: cBAD
 - Beste Lösung kommt von der Uni Rostock, Citlab Team

- 8) wegen Vertretung über die Güter...
- 9) Die aus der Freistellung dieses Vertrages, sowie aus dem...
 lichen Einverständnis des Falles von dem...
 aus dem die Vertretung...
 aus der...
 10) für alle aus diesen...
 Praxistexten...
 für alle aus diesen...
 Praxistexten...
 für alle aus diesen...
 Praxistexten...
- 11) Die Vertretung wird in einem Exemplare, welches in...
 der k.k. Staatsbahn-Direktion Innsbruck...
 aufbewahrt und erfüllt durch Josef Schipfmair, einen...
 Abkömmling dieser Urkunde.

Witzbühel, am 5. August 1910
Innsbruck, am 24. Oktober 1910
 k.k. Staatsbahn-Direktion Innsbruck
 Dr. k.k. Staatsbahn-Direktion
 L. B. Wagner

72k 111
 dem...
 Aufschneider
 diese...
 von der...
 gerichtliches Witzbühel,
 am 5. August 1910

Der mit... Urkunden und
 Legalisierungstempel versehenen Urkunden
 gleich
 K. k. Bezirksgericht Kitzbühel
 am 17. Jänner 1911



- 8) wegen Vertretung über die Güter...
- 9) Die aus der Freistellung dieses Vertrages, sowie aus dem...
 lichen Einverständnis des Falles von dem...
 aus dem die Vertretung...
 aus der...
 10) für alle aus diesen...
 Praxistexten...
 für alle aus diesen...
 Praxistexten...
 für alle aus diesen...
 Praxistexten...
- 11) Die Vertretung wird in einem Exemplare, welches in...
 der k.k. Staatsbahn-Direktion Innsbruck...
 aufbewahrt und erfüllt durch Josef Schipfmair, einen...
 Abkömmling dieser Urkunde.

Witzbühel, am 5. August 1910
Innsbruck, am 24. Oktober 1910
 k.k. Staatsbahn-Direktion Innsbruck
 Dr. k.k. Staatsbahn-Direktion
 L. B. Wagner

72k 111
 dem...
 Aufschneider
 diese...
 von der...
 gerichtliches Witzbühel,
 am 5. August 1910

Der mit... Urkunden und
 Legalisierungstempel versehenen Urkunden
 gleich
 K. k. Bezirksgericht Kitzbühel
 am 17. Jänner 1911



in lingua... ~~...~~

de... ~~...~~

De... ~~...~~

~~...~~

...

Ad... ~~...~~

...

De... ~~...~~

~~...~~

...

~~...~~

...

...

Province de l'équinoxiale

OBSERVATIONS MÉTÉOROLOGIQUES

Territoire de l'équinoxiale

District de la Eschurpas

du mois de Janvier 1935 (faites à 6 heures.)

Poste AM Jardin Botanique d'Etat

Observateur J. Denis

Date	BAROMÈTRE			THERM. S.A.		PSYCHROMÈTRE			ACTINOMÈTRE		GIRQUETTE			NUAGES		EAU TOMBÉE
	TEMP.	HAUTEUR LU	HAUTEUR QAG	MINIMA	EVAP.	SEC	HUMIDE	HUMIDITÉ	☉	☾	DIRECTION	FORCES	NEB.	FORMES		
1	26.1	737.0	734.2	24.5	1.0	22.9	22.4	21	21.2	22.7	Variable	0	3	cc	-	
2	26.3	736.9	734.1	24.7	1.1	21.3	21.0	21	20.5	20.6	N	1	6	cc	-	
3	26.1	736.7	734.3	24.8	1.1	21.1	20.7	21	20.1	20.2	N	1	9	cc	-	
4	26.0	736.3	734.2	24.8	0.9	22.0	21.4	25	21.5	21.6	E	1	7	cc	-	
5	25.0	737.0	734.2	24.5	1.1	22.0	21.6	24	22.4	22.5	S	1	10	cc	-	
6	25.0	737.0	734.3	24.6	1.0	22.3	21.9	26	21.3	21.4	Variable	0	7	cc	6.0	
7	25.2	736.9	734.5	24.8	0.4	22.3	21.1	25	21.1	21.3	N	2	10	cc	-	
8	24.7	736.9	734.5	24.7	0.8	22.5	22.0	25	22.2	22.1	N	1	3	cc	-	
9	24.0	737.0	734.3	24.4	1.2	22.2	21.8	26	21.6	21.7	Variable	0	2	cc	-	
10	24.8	736.9	734.2	24.5	0.9	21.9	21.4	26	21.9	22.1	S	1	6	cc	22.0	
11	24.7	737.1	734.6	24.5	1.4	21.2	20.8	26	20.3	20.4	N	1	8	cc	-	
12	24.9	736.8	734.7	24.8	0.9	21.1	21.4	24	21.1	21.3	N.E.	1	2	cc	-	
13	24.1	736.9	734.2	24.0	0.7	22.2	21.8	22	21.5	21.6	E	1	6	cc	-	
14	24.7	737.0	734.2	24.1	1.1	22.1	21.7	21	22.1	22.2	Variable	0	7	cc	2.1	
15	24.8	736.4	734.7	24.7	1.3	21.5	21.2	24	20.7	20.8	N	3	10	cc	1.4	
16	24.6	737.0	734.3	24.5	1.0	22.3	22.0	25	21.8	21.9	E	3	9	cc	-	
17	24.6	737.0	734.3	24.5	1.3	22.1	21.3	25	22.1	22.3	N	1	2	cc	-	
18	24.7	736.9	734.2	24.8	1.2	22.0	21.5	25	21.4	21.5	N	2	6	cc	26.0	
19	24.0	736.9	734.4	24.8	1.5	21.1	21.2	23	20.6	20.8	N	2	10	cc	-	
20	24.1	737.0	734.3	24.9	0.9	21.9	21.5	24	21.3	21.5	S.E.	1	8	cc	-	
21	23.0	736.6	734.1	24.7	0.7	22.0	21.6	26	21.2	21.1	N.E.	1	7	cc	-	
22	24.1	736.9	734.1	24.5	1.2	22.0	21.5	26	22.3	22.4	N.N.W.	1	7	cc	-	
23	24.9	736.9	734.1	24.3	1.3	22.0	21.7	27	22.5	22.2	N	1	2	cc	0.2	
24	24.0	736.9	734.1	24.7	0.1	22.3	22.0	26	21.8	21.7	S.E.	1	8	cc	-	
25	24.3	736.7	734.1	24.3	0.6	22.3	21.9	26	21.5	21.6	S.E.	1	7	cc	-	
26	24.8	736.7	734.4	24.5	0.9	22.3	21.1	25	21.9	21.0	N	1	7	cc	-	
27	24.3	736.9	734.2	24.4	0.8	21.6	21.1	26	22.1	22.3	N.N.W.	1	7	cc	-	
28	24.2	737.0	734.2	24.6	1.7	22.4	21.9	26	22.7	22.6	Variable	0	7	cc	0.6	
29	24.5	737.0	734.2	24.6	1.5	22.3	21.3	25	22.1	22.3	N.N.W.	1	2	cc	-	
30	24.7	737.0	734.7	24.6	1.3	22.2	21.9	24	22.5	22.1	E	1	6	cc	1.4	
31	24.1	737.0	734.1	24.7	1.1	22.6	22.1	24	22.3	22.5	S.E.	1	2	cc	-	

Moyennes	24.1	736.1	734.1	24.9	3.3	22.3	21.8	25	21.5	21.6			41			7.4
MAX ET MIN ABSOLUS DE T°																12.1
FREQUENCE DES VENTS																LE 7
																LE 7
																TOTAL EAU TOMBÉE
																71.7 mm
																NOMBRE DE JOURS DE PLUIE
																1
																DE BROYILLARD
																28 TONNERS
																28 GRÊLE

Resultate

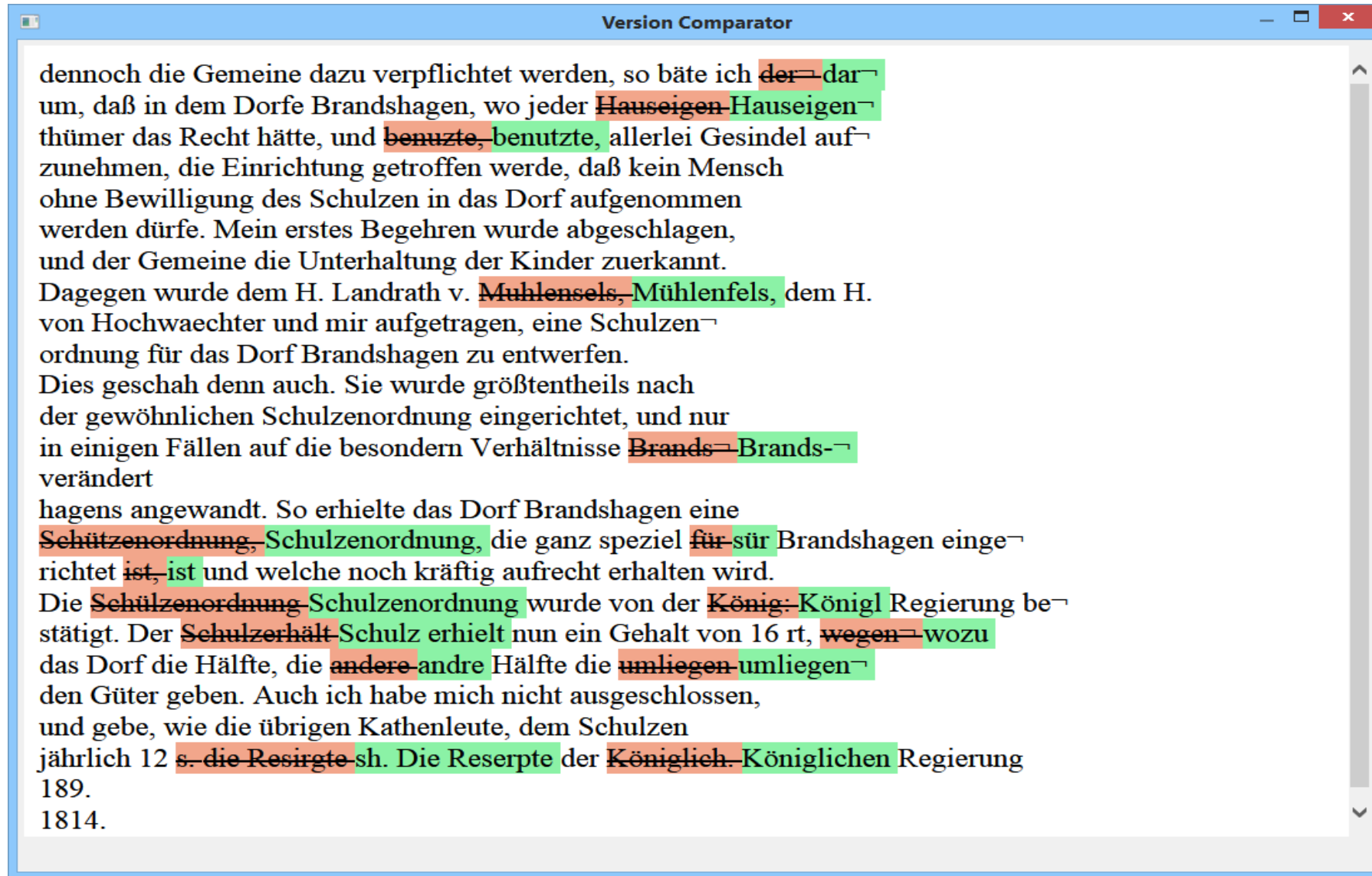
Konzilsprotokolle

- Universität Greifswald, Beschlüsse
- Spätes 18. Jahrhundert, deutsche Kurrentschrift
- Ein Schreiber
- Trainingset: 35.743 Wörter = 182 Seiten
- Character Error Rate (CER) am Testset = 3,1% (ohne Wörterbuch)
- Word Error Rate (WER) am Testset = 13,1% (ohne Wörterbuch)

Innen die Gemeine dazu veranlaßt worden, so bitten wir Sie,
um, daß in dem Dorfe Braudsfagen, wo jedes Haus einen
Gärtchen das Recht hätte, und benutzt, allenthalben gesündel auf-
zuwachen, die Einrichtung getroffen werden, daß die Meist-
offen Lössilligung das Dorf in das Dorf aufzunehmen
werden dürfen. Mein nebst Lössilligung werden abgefahren,
und die Gemeine die Unterhaltung der Lössilligung zuordnen.
Dagegen werden dem H. Landrat v. Mühlentfels, dem H.
von Hochwächter und mir aufzutragen, eine Besondere
Ordnung für das Dorf Braudsfagen zu unternehmen.

189.

1814.



Für diese Seite: CER = 2,2% / WER = 10,3% (mit Wörterbuch)

Mittelalterliche Schriften

- Kooperation mit Dominique Stuetzmann und CNRS (Institut de recherche et d'histoire des textes) Paris
- HIMANIS Projekt
- Viele verschiedene Schreiber
- Französisch und Latein
- Trainingset: 550.381 Wörter oder 1197 Seiten
- CER am Testset = 6,4%
- WER am Testset = 22,1%

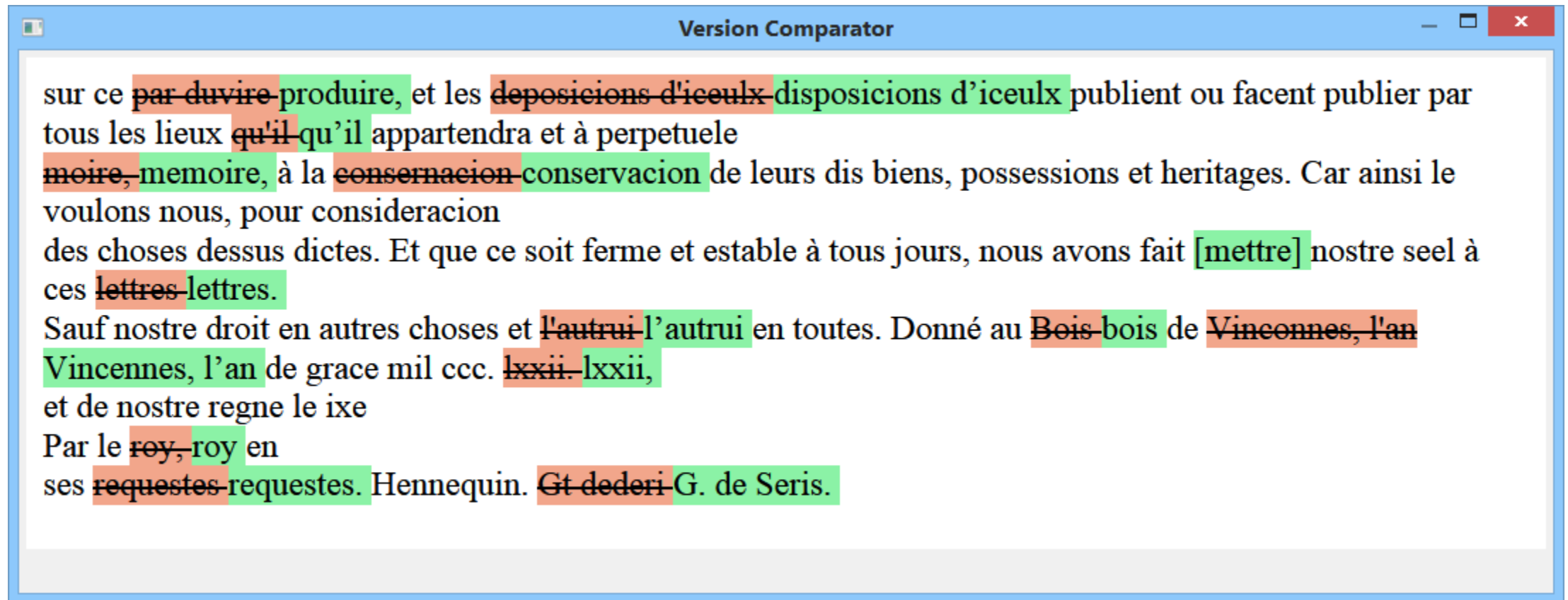
en a l'empereur et a chascun d'eux en son royaume si n'est ce q'ou car. Et de ce q'ou met le dit Anulle en possession et pais
 de la terre et possessions deff dres Royumes et deff dres Justis ala Galem deff dres l'un d'eux de rente deff dres en la susdit
 layse son et d'yr perpetuel et paisiblement et ses heirs et heirs q' de la amont car d'icelle et de ses appartenances
 depend de de p'ngue hieute et avecq le f'ent d'icelle q' les p'ngue a q'il appartient a son et h'omage de la d'icelle
 et possessions deff dres se messent a celle sul p'ant ou red'icelle car pom car d'icelle de d'icelle et de d'icelle q' h'omage
 none en soit du le d'icelle ala f'ant et le d'icelle de d'icelle n' morte q' d'icelle none pom none en f'ant la f'ant
 et h'omage n'ng n'ndit en emp' d'icelle aucun car d'icelle de d'icelle n'ndit fait de d'icelle d'icelle d'icelle d'icelle
 obstant aucun d'icelle p'none f'ant d'icelle d'icelle et d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle
 sub'icte as d'icelle d'icelle Et poir et d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle d'icelle
 in l'oy et d'icelle d'icelle le d'icelle en mois d'icelle
 l' de f'ant

par le Roy empereur
 J. de...

Visa

estre et tan sei estoient retourne en la dite man
et y les freres du dit lieu et recelle longuement
res sestoient consentis en donnant aus forfaissans
les forfaissans dessus dis et en leur donnant
ave estre dessus dis seimes appeler en droit y d
longuement les freres de la dite man

fin ce qu'il me et les Supplémentes de
monie a la confirmation de l'œuvre de
des choses de plus de. Et que
Gauf me soit en ces choses
et de mes tégnes de un



Für diese Seite: CER = 6,02 / WER = 19,6 (ohne Wörterbuch)

Gedruckter Text - Zeitungen

- Wiener Diarium – in Kooperation mit Österreichischer Akademie der Wissenschaften
 - Zeitung aus dem 18. Jhd.
 - Bitonale Scans
 - Trainingset: 179.997 Wörter oder 345 Seiten
 - CER am Testset = 0,81
 - WER am Testset = 3,02
- Vergleichbare Ergebnisse mit NZZ Modell – für ca. 160 Jahre (1780 – 1940 1 Titelseite pro Jahr – insg. 167 Titelseiten) – CER am Testset: 0,47% – impresso Projekt

Entbindung herannahete, so wurden bestwe-
gen sogleich bey Hof alle nothwendige Veran-
staltungen vorgelehret, auch darumen sowohl in
denen Kaiserl. Hof. Kapellen, als in denen
alhieyigen Pfarz. und mehreren Kirchen das
Hochwürdigste Altars. Sacrament unter Zu-
lauf deren alhieyigen getreuesten Inwohnern
zu Erbitung einer höchst. beglückten Entbin-
dung für Ihre Kaiserl. Königl. Majestät aus-
gesetzt: auch hatte ein zahlreicher hoher Adel
nach Hofe sich erhoben, worauf sodann um 9.
Uhr Abends unsere Allermüdeste Landes. Mut-
ter mit einer schönen und gesunden Erz. Her-
zogin, Gott lob, glücklich zu allgemeiner Freu-
de deren unterthänigsten getreuen Inwohnern
alhieyiger Kaiserl. Königl. Haupt. und Resi-
denz. Stadt Wien, entbunden worden: auch
seynd viele Botten zu Pferd und zu Fuß von
hier abgeschicket worden, diese glückliche Zei-
tung auch auf dem Land kund zu machen. Fol-

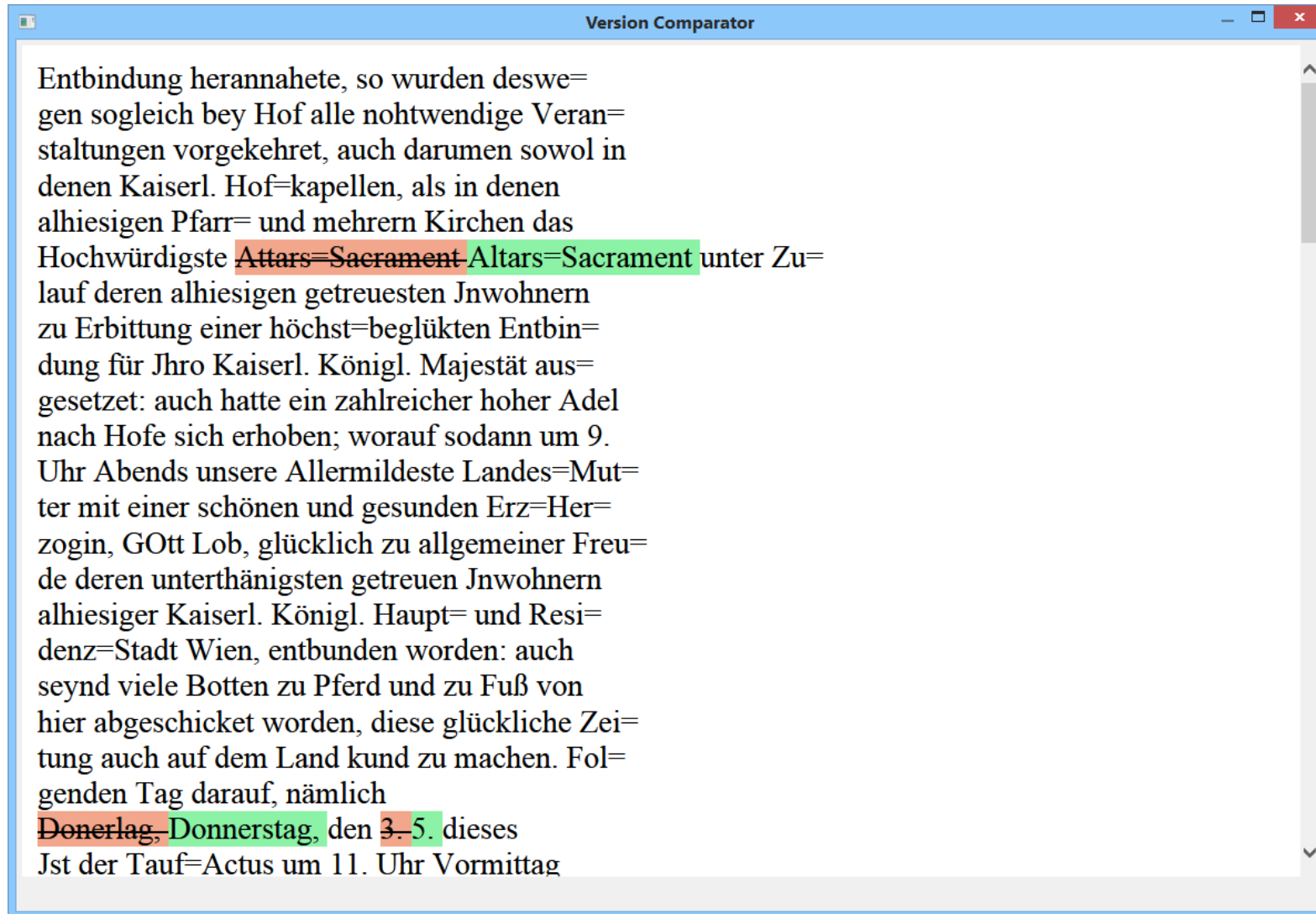
Kaisers, der gesamten Kaiserl. Königl. jungen
Herrschaft, und des in größter Anzahl und
prächtiger Gala, welche auch durch die 2. fol-
gende Tagen gedauert, erschienenen hohen
Adels verrichtet.

Mittags haben Allerhöchst. besagt Se. Ma-
jestät der Kaiser unter der gewöhnlichen Auf-
wartung des Päpfl. Hrn. Nuntii, und des
zahlreichen hohen Adels öffentlich gespeiset.
Abends haben Se. Excell. (Tit.) der Herr
Vio Graf von Bathiani denen Durchlaßten
jungen Herrschaften samt einigen jungen Da-
mes und Cavalieren in seiner Wohnung bey
Hof einen Ball und prächtiges Abendmahl
gegeben. Die Theatra nächst der Kaiserl. Burg
und dem Kärntner. Thor waren des Abends
prächtig illuminiert, und seynd alle Stadt-
Thore Sperr. frey gewesen.

Freitag, den 6. dieses

Nach dem gewöhnlichen Gottes. dienst in
den Kirchen hatten alle eine Messe

Mittags haben Allerhöchst. befagt Se. Ma-
jestät der Kaiser unter der gewöhnlichen Auf-
wartung des Päpstl. S:n. Nuntii, und des



Für diese Seite: CER = 0,6 / WER = 3,0% (ohne Wörterbuch)

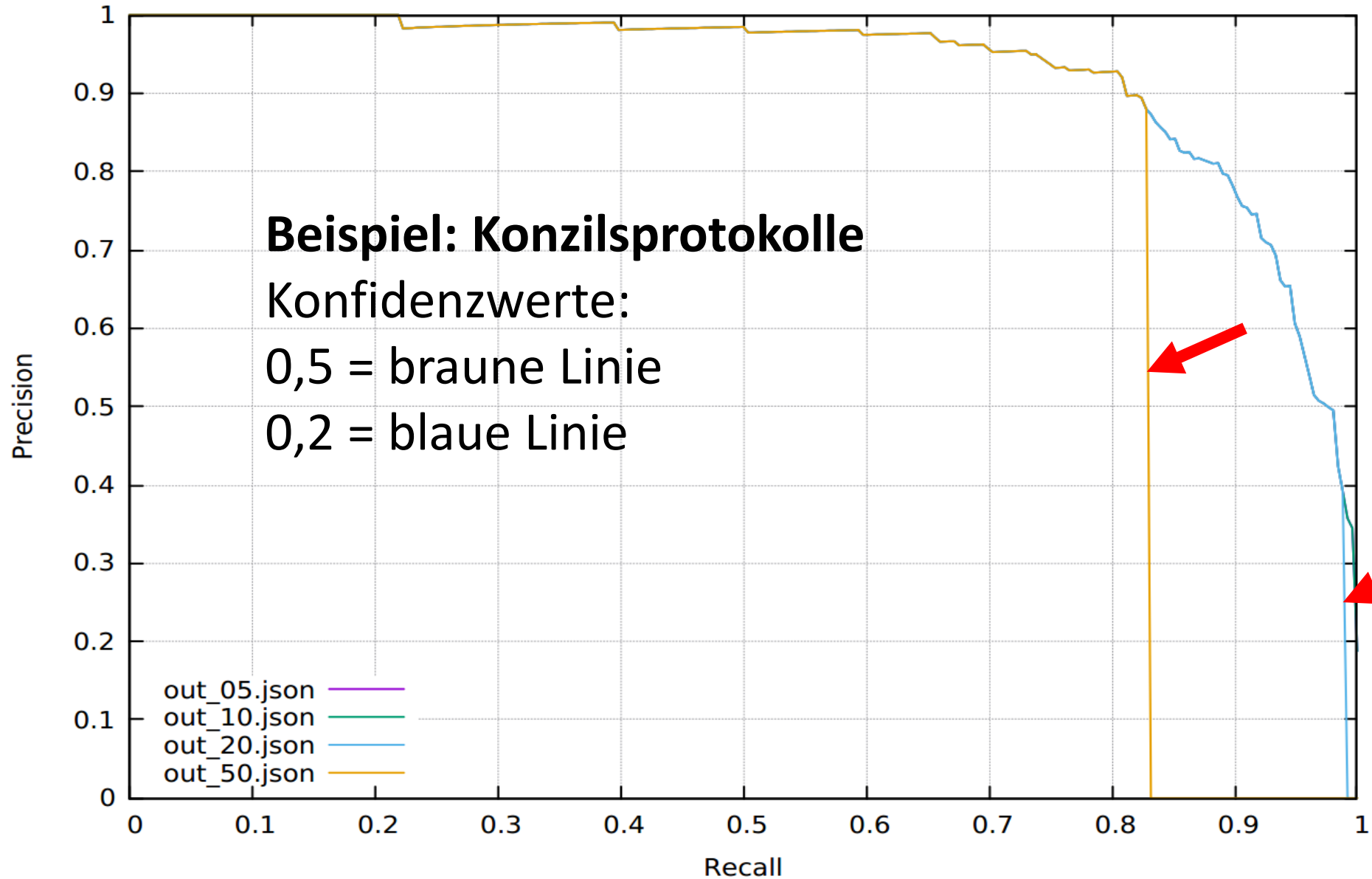
Layout Analyse und automatisierte Texterkennung für historische Dokumente zeigen exzellente Resultate für gedruckte Schriften und gute bis sehr gute Ergebnisse für handschriftliche Dokumente.

Keyword Spotting

Keyword Spotting (KWS)

- Hintergrund
 - Neuronale Netze liefern mehr als nur den reinen Text
 - Sogenannte Konfidenztabelle
- Keyword Spotting
 - Eine effiziente Methode um diese Konfidenztabelle für die Suche nutzbar zu machen
 - Der Benutzer kann selbst entscheiden wie tief er gehen möchte
 - Geht es darum rasch “etwas” zu finden – oder ein bestimmtes Wort – z.B. einen Familiennamen – sicher in einem Dokument zu entdecken

Precision-Recall-Curve



Search for...

Documents | Fulltext (Solr) | Tags | **KWS**

Search in: Current document Partial Matches Case-sensitivity Expert Syntax

Confidence Threshold: < >

Queries

Keyword 1

Keyword 2 X

Keyword 3 X

Keyword 4 X

Keyword 5

Search

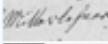
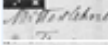
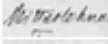
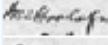
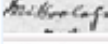

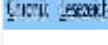
Search Results

Created	Status	Queries	Duration	Scope
23.01.18 18:13:18	Completed	"MITTERLEHNER" (23), "FRANZ" (30), "MA...	4.54 sec.	Document 300...
29.12.17 13:22:14	Completed	"Mitterlehner" (17), "Thening" (4)	3.34 sec.	Document 300...
05.12.17 16:21:21	Completed	"Linz-Waldegg" (0)	1.15 sec.	Document 300...
23.11.17 14:26:20	Completed	"Thening" (4)	4.02 sec.	Document 300...
23.11.17 12:49:47	Completed	"Mitterlehner" (14)	1.39 sec.	Document 300...
23.11.17 12:47:49	Completed	"Krankenhaus" (3), "Thening" (16)	2.02 sec.	Document 300...
23.11.17 12:45:30	Completed	"Franz Mitterlehner" (3), "Franz" (5), "Mitter...	3.23 sec.	Document 300...

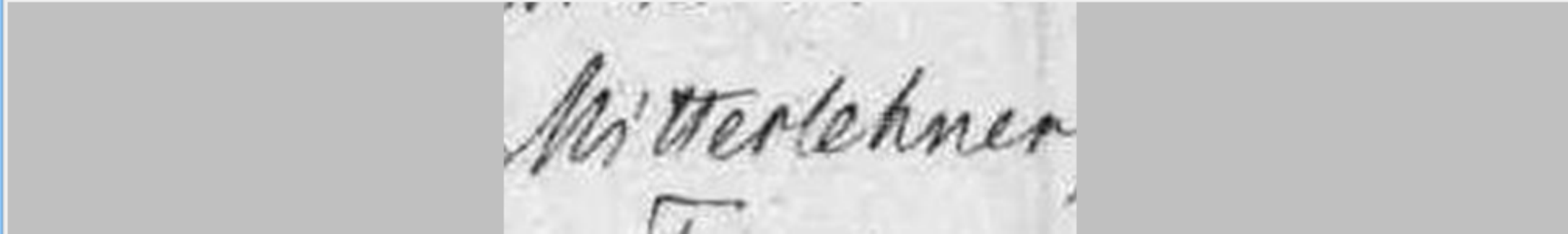
Close

Keyword Spotting Results

"Franz Mitterlehner" (3 hits) "Franz" (5 hits) "Mitterlehner" (7 hits) "Linimeir" (2 hits)

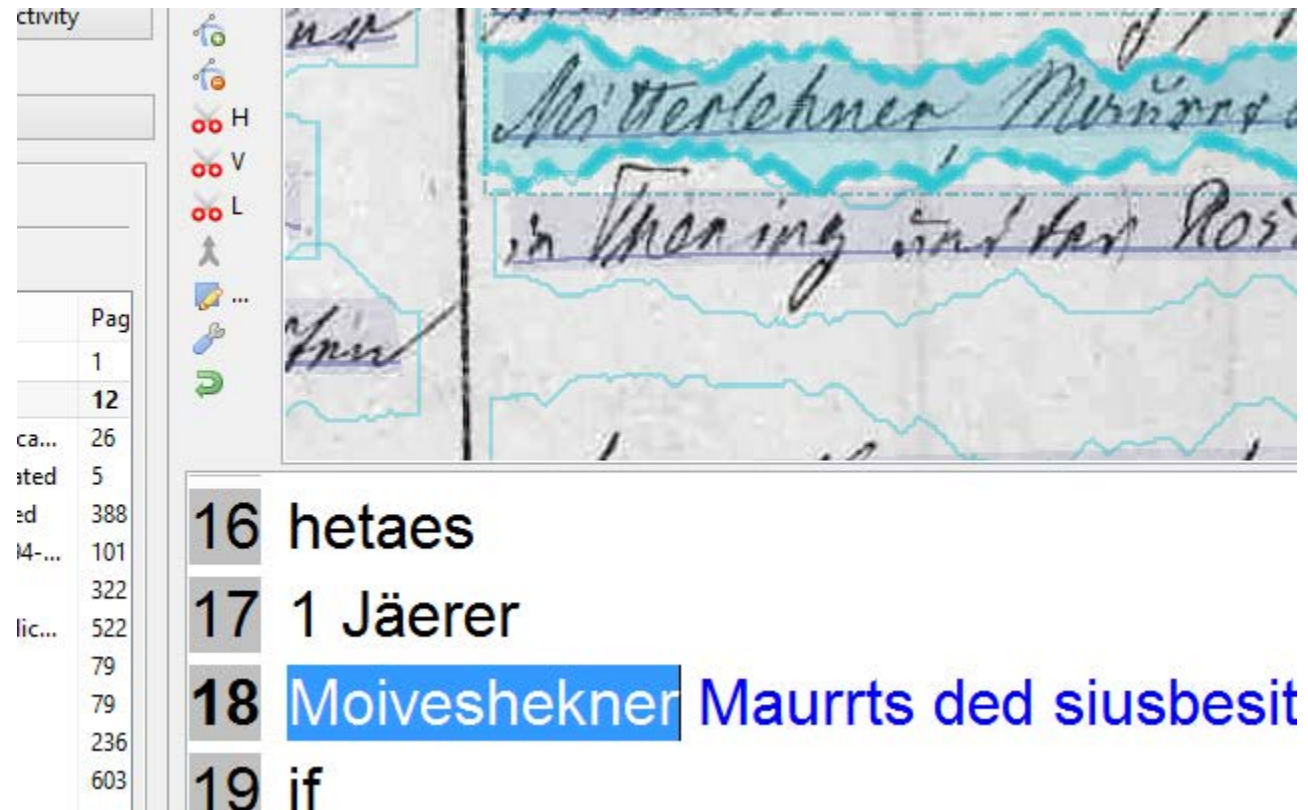
Confidence	Page Nr.	Line transcription	Previ...
0.4983	3	des Mitherlehaer	
0.4707	1	dranz Miterlekner, dienskaachtsein	
0.2080	1	Moiveshekner Maurrts ded siusbesitzas	
0.1445	4	Jofuch M strlofet	
0.1429	5	Josiuch Mberlehent	
0.0574	1	z e Mensrgaaeeghdnetuhes. Msmernensgriduach Mti	
0.0530	4	t keeste seht Mnet lesnnehen Eni Me	

Preview



Close

Mitterlehner - Moiveshekner



The image shows a digital document viewer interface. The main area displays a handwritten manuscript page with blue wavy lines highlighting the text. The text is written in a cursive script and includes the words "Mitterlehner" and "Moiveshekner". To the left of the manuscript is a sidebar with various icons and a table of contents. The table of contents lists page numbers and corresponding text entries.

Page	Text
1	
12	
26	ca...
5	ited
388	ed
101	14-...
322	
522	lic...
79	
79	
236	
603	

16 hetaes

17 1 Jäerer

18 Moiveshekner Maurrts ded siusbesit

19 if

Bentham Papers Indexing and Search

Bentham papers

193 Boxes (89,911 indexed page images)

Total number of spots:	197,651,336
Estimated num. of running words:	25,487,932
Estimated num. of run. words per page:	283
Average num. of spots per page:	2,198
Average num. of spots per run. word:	7.8

Bentham Papers Indexing and Search

PRHLT engine **BETA**

Austria

PRHLT READ UCL

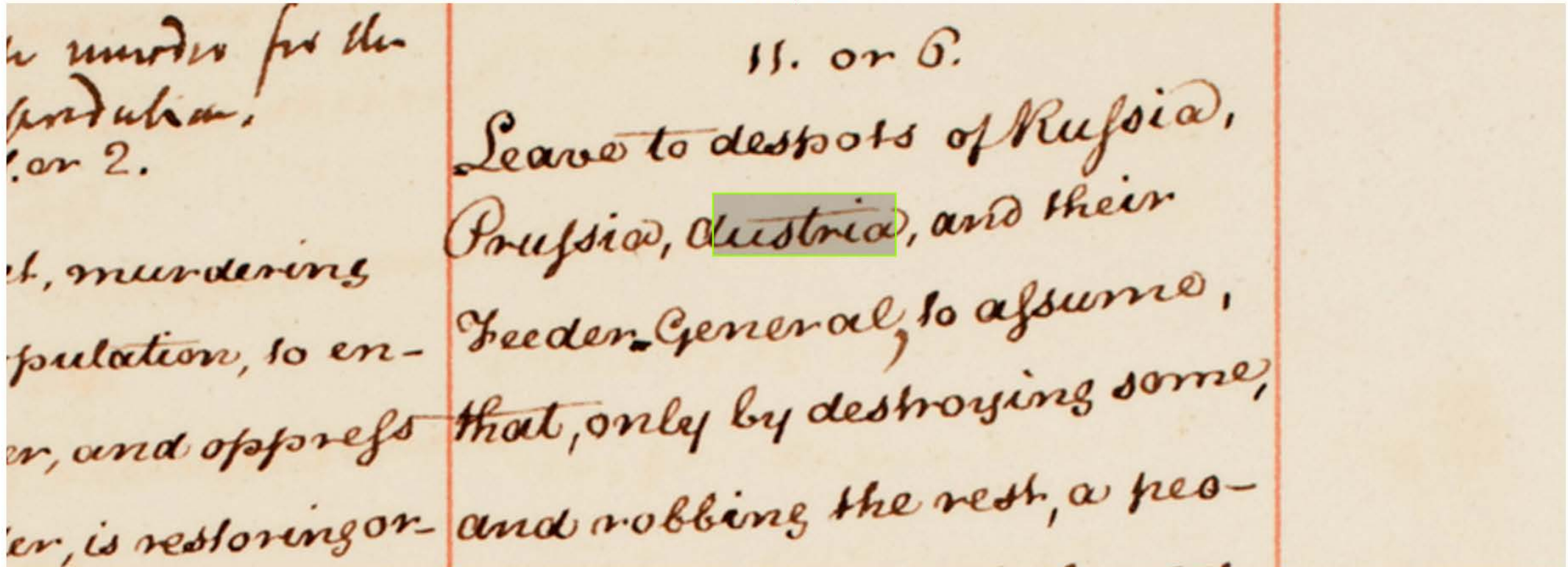
Help

Confidence: 50 Max. results:

You are here: [home](#) » [BENTHAM](#) » [Box008](#) » [page 97 \(008_088_001\)](#) **Hand:** John Flowerdew Colls

1 match found for "austria" with a confidence of 71.2%

[← Previous](#) | [Next →](#)



Austria

Help

Confidence: 50



Max. results:



You are here: [home](#) » [BENTHAM](#) » [Box008](#) » [page 97 \(008_088_001\)](#)

Hand: John Flowerdew Colls

1 match found for "austria" with a confidence of 71.2%

[← Previous](#) | [Next →](#)

Transkribus – Expert client

Server Overview Layout Metadata Tools

Logout guenter

Document Manager User Manager

Versions Jobs

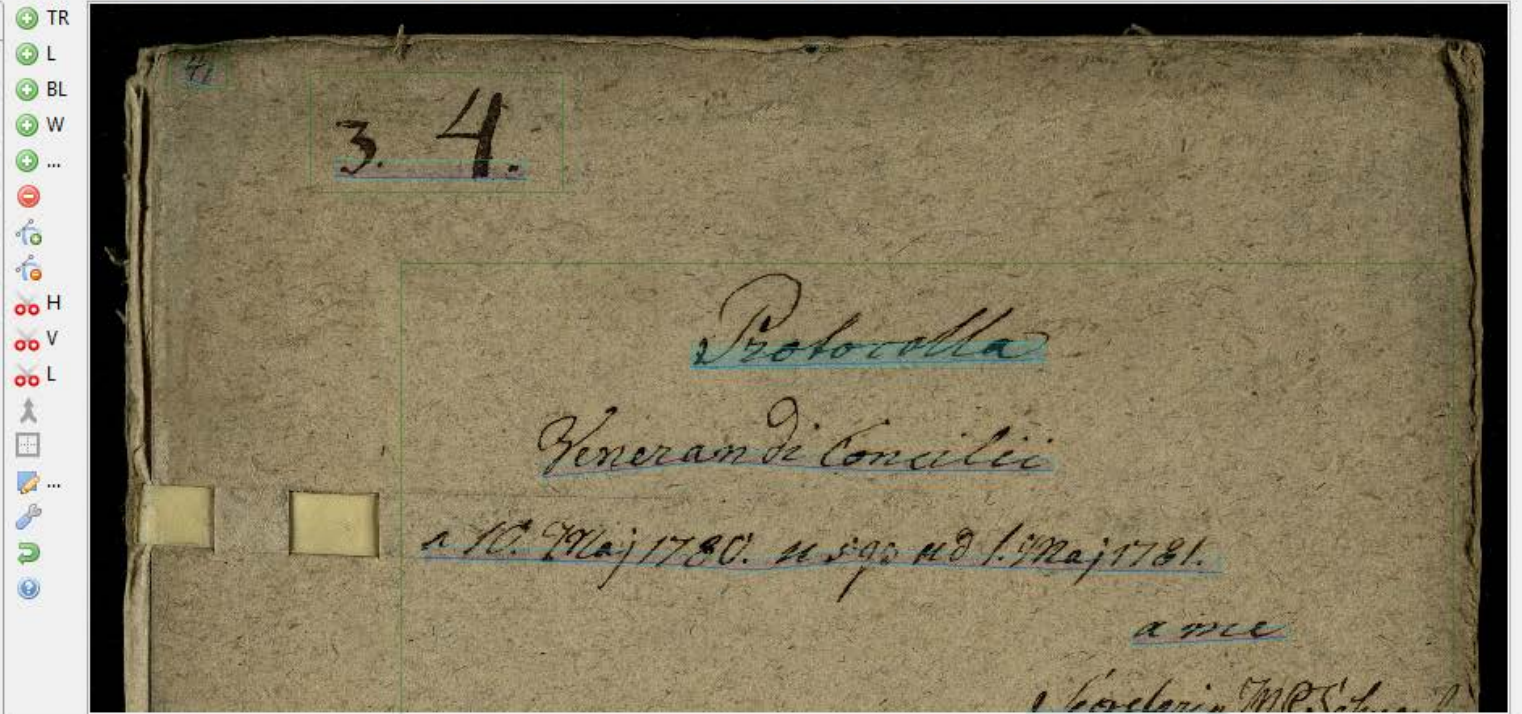
Recent Documents... User activity

Collections: eLearning (2805, Owner)

1-49 / 49 1 1

ID	Title	Pages	Uploa
23707	\\ad.ucl.ac.uk\homeE\uczwlse\Deskto...	19	louise
18545	Bozner Ratsprotokolle (Hs. 115), 18. Jh.	27	barba
18054	Dutch notarial paper (A28918)	20	guent
18053	Kochbuch (Sütterlin), 1926	11	guent
17188	Bozen Ratsprotokolle HS_115	11	guent
8583	Barock_Greifswald_1720	6	alvern
8569	Barock_Greifswald_1696	4	alvern
8568	Letter from Greifswald	6	alvern
8567	Renaissance_Greifswald_1564_Abbrev_...	6	alvern
8557	StAZH: Protokolle des Regierungsrats ...	10	tobias
8556	StAZH: Protokolle des Kleinen Rats 1807	10	tobias
7048	GTset_ABP_selection_200_II	85	eva.la
6454	Jeremy Bentham 1800	10	guent
6323	Zwettl 1653	112	guent
6322	Reichsgericht 1904	16	guent
6320	Konzilsprotokolle 1794/55	31	guent

100 Filter



1-1 4/
2-1 3. 4.
3-1 Protocolla
3-2 Venerandi Concilii
3-3 a 10. Maj 1780. usque ad 1. Maj 1781.
3-4 a me

HTR Training

Model Name:

Language:

Description:

CITlab HTR CITlab HTR+ CITlab T2I

Nr. of Epochs:

Training Set

- ▶ 142992 - TRAINING_TESTSET_Amerikaauswanderer_M1+_duplicated (▲
- ▶ 141855 - mist2 (10 pages)
- ▶ 141798 - mist (10 pages)
- ▶ **138225 - TEST_CITlab_AHB_19-03-14_duplicated (4 pages)**
- ▶ 137698 - NAFCourtRecordsSample_duplicated (100 pages)
- ▶ 135987 - TEST_CITlab_MS A2654 M1+_duplicated (2 pages)
- ▶ 135424 - TEST_CITlab_Kepler Somnium M2_duplicated (3 pages)
- ▶ 134871 - diplomatische lage (6 pages)
- ▶ 132141 - TEST_CITlab_Noscemus M1_duplicated (11 pages)
- ▶ 131608 - Sample document 501594 (10 pages)
- ▶ 131606 - Sample document 501591 (4 pages)
- ▶ 131605 - Sample document 501589 (100 pages)
- ▶ 131005 - TEST_CITlab_Gusbeth M3_duplicated (1 pages)
- ▶ 130991 - TRAIN_CITlab_Széchényi German Handwriting (1 pages)

Overview

- Use Groundtruth versions
- Use initial('New') versions

Training Set

ID	Title	Pages
138225	TEST_CITlab_AHB_19-03-14_duplicated	1-4

Test Set

ID	Title	Pages
135424	TEST_CITlab_Kepler Somnium M2_duplicat...	1-3

Choose a model

Search: Technology: Show all ▾

Name	Language	Curator
Frans HW 10.0	french	rpsimmelink@
JohannAdam-4	German	hj.oertel@t-or
AvS Relation	German, Latin	aqschweizer@
TUNI_kuolleidenluetteloTUIF2_TUIF3_f...	Finnish	oskari.jokineni
BvgA BNU	German	antje.willing@
TUNI_kuolleidenluetteloTUIF2_TUIF3_f...	Finnish	oskari.jokineni
Sütterlin Handschrift	German	brehm79@goc
Referat fra regjeringskonferanser Mixe...	Norsk	stejac@arkivve
TB Mr 4	Latein	plovertc@gma
NA_Daghregister_original_M1+	dutch	guenter.hackl
French HW 2.1	french	rpsimmelink@
Regjeringskonferanser Handwriting In...	norsk	stejac@arkivve
Regjeringskonferanser Handwriting In...	Norsk	stejac@arkivve
Letter Family terpinc 9000, 50 epochs	german	joze.glavic1@q
Wittgenstein_HTR	German	florian.landes@
Regjeringskonferanser Handwriting In...	Norsk	stejac@arkivve
Regjeringskonferanser Handwriting In...	Norsk	stejac@arkivve
test acceptatie omgeving	Dutch	vincent.noppe
FinnishThesis	Finnish	mari.siiroinen@
Konzepte M_1	german latin	alverman@un
Kyryll_11_Century_full	Russian Churc...	achim.rabus@
Referat fra regjeringskonferanser Mixe...	Norsk	stejac@arkivve
Notary De Winter, 610 scans	Dutch	l.van.der.werff
V4	French	antoinedesacy

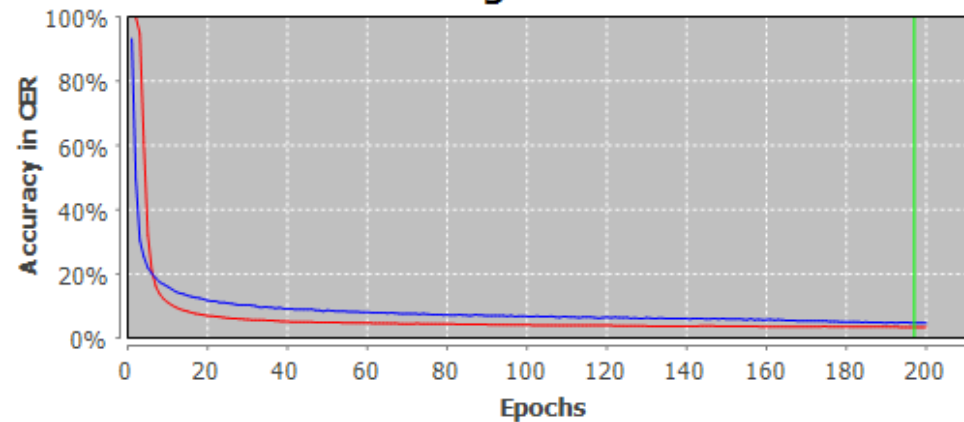
Details

Name: Language:

Description: Parameters:

Nr. of Words: Nr. of Lines:

Learning Curve



CER on Train Set: CER on Test Set:

OK

Cancel

Advanced Statistics

Page	WER	CER	Word Acc	Char Acc	Bag Tokens Prec	BT Recall	BT F1-Score
Overall	29.25 %	8.97 %	70.75 %	91.03 %	0.752	0.746	0.752

Page	WER	CER	Word Acc	Char Acc	Bag Tokens Prec	BT Recall	BT F1-Score
Page 4	22.42 %	5.52 %	77.58 %	94.48 %	0.8116	0.8136	0.81257861635...
Page 5	19.06 %	5.22 %	80.94 %	94.78 %	0.8421	0.85	0.84603421461...
Page 6	49.52 %	13.62 %	50.48 %	86.38 %	0.6077	0.6208	0.61420118343...
Page 7	18.55 %	4.66 %	81.45 %	95.34 %	0.8369	0.85159999999...	0.84414327202...
Page 8	30.0 %	8.44 %	70.0 %	91.56 %	0.7225	0.741	0.73164556962...
Page 9	28.34 %	7.15 %	71.66 %	92.85 %	0.75959999999...	0.7664	0.76297968397...
Page 10	50.70 %	15.28 %	49.30 %	77.78 %	0.4640	0.4720	0.46026406026...

Compare Text Versions for Page 4

Error Rate Chart | Ref: GT | Hyp: CITlab HTR: NAF2019 Court Records M1

■ WER
 ■ CER

Show all results

Base folder:

File/Folder name:

Export path:

Download XLS

Server Overview Layout Metadata Tools

Layout Analysis

Method: CITlab Advanced

Current page Pages (26) 1-26

Find Text Regions

Find Lines in Text Regions

Text Recognition

Method: HTR (CITlab)

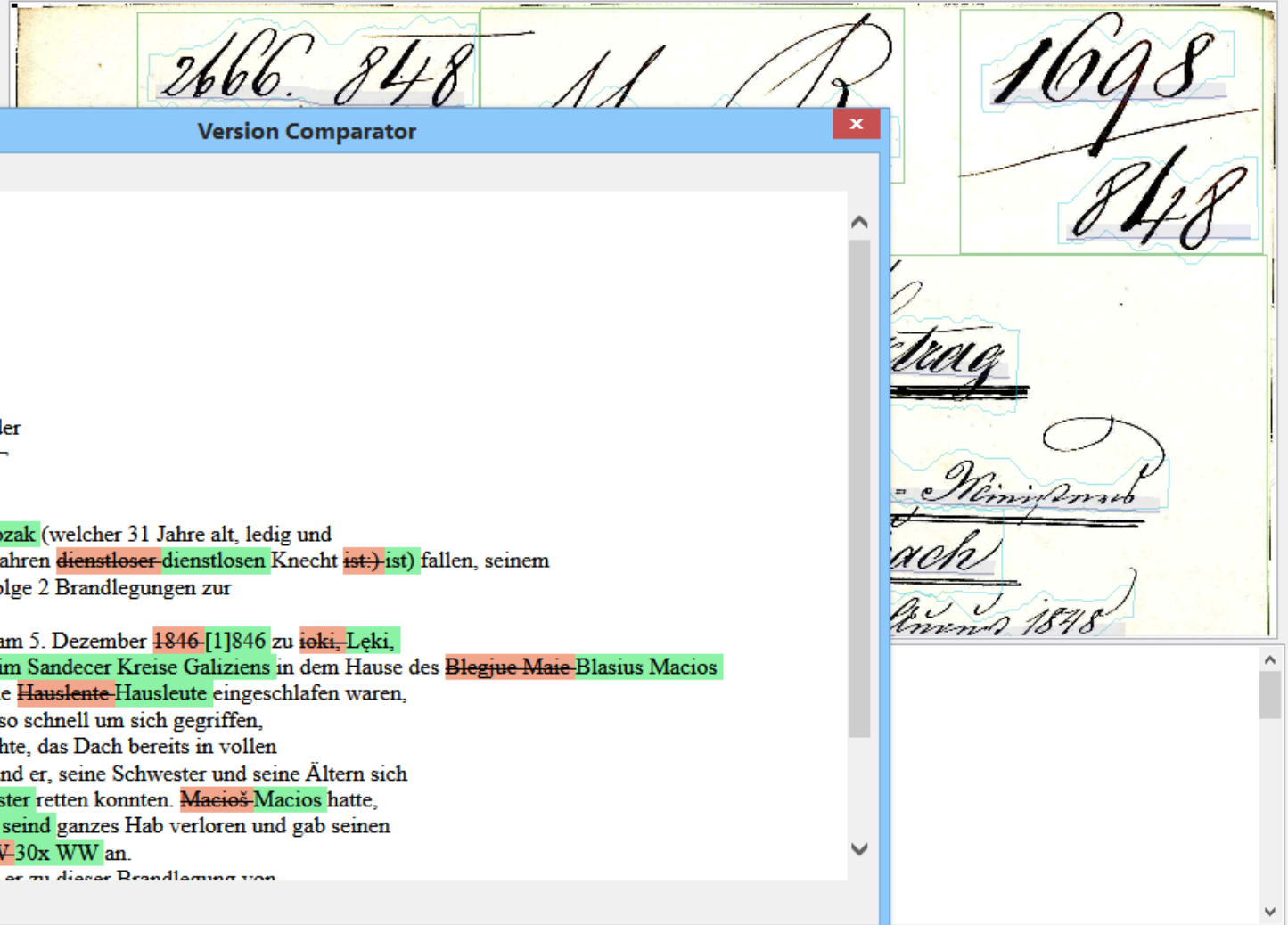
Compute Accuracy

Reference: (Correct Text) 31.01.19 11:16:29 - guenter - Gro

Hypothesis: (HTR Text) 31.01.19 11:17:21 - guenter - In Pro

Other Tools

Current page Pages (26) 1-26



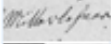
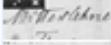
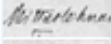
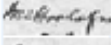
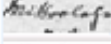
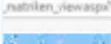
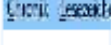
Version Comparator

Show line numbers

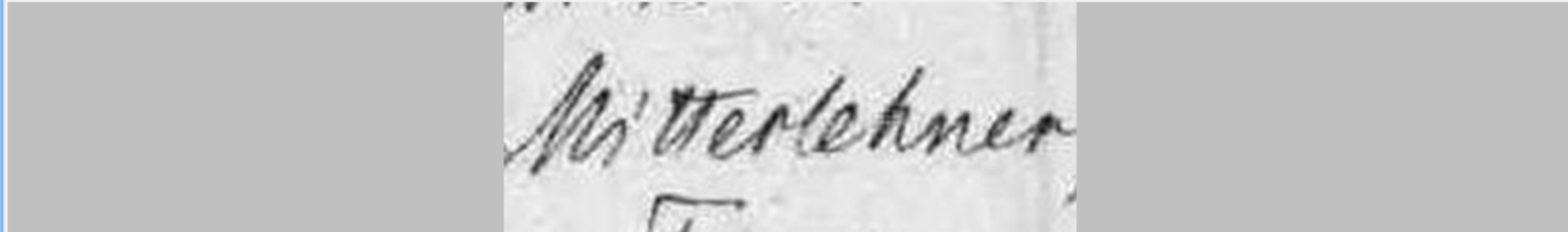
2666-848-2666 [1]848
M.A-MR
1698
848
Vortrag
des Justiz-Ministers
~~Id. ach~~ Dr. Bach
vom 4. August 1848
mit dem Todesurtheile wider
Peter Kozak wegen Brand-
legung und dem Antrage
auf Begnadigung
Dem ~~Peter Kozak~~ Peter Kozak (welcher 31 Jahre alt, ledig und
~~einschon ein~~ schon seit 2 Jahren ~~dienstloser dienstlosen~~ Knecht ~~ist:-) ist~~) fallen, seinem
eigenen Geständniße zu Folge 2 Brandlegungen zur
~~Ksht~~ Last.
Den ersten Brand hatte er am 5. Dezember 1846-[1]846 zu ~~loki~~ Leki,
~~di Soderen Krece belzend im Sandecr Kreise Galiziens~~ in dem Hause des ~~Blegiue Maie~~ Blasius Macios
Nachts, nachdem schon alle ~~Hauslente~~ Hausleute eingeschlafen waren,
gelegt. Dieses Feuer hatte so schnell um sich gegriffen,
~~daß, daß~~ als Macios erwachte, das Dach bereits in vollen
~~flammen~~ Flammen stand und er, seine Schwester und seine Ältern sich
nur durch das ~~Fenster~~ Fenster retten konnten. ~~Macios~~ Macios hatte,
außer einer ~~Kuh sein~~ Kuh, ~~seind~~ ganzes Hab verloren und gab seinen
Schaden auf 150f ~~305 M W~~ 30x WW an.
Der Inquisit sagte aus, daß er zu dieser Brandlegung von

Keyword Spotting Results

"Franz Mitterlehner" (3 hits) "Franz" (5 hits) "Mitterlehner" (7 hits) "Linimeir" (2 hits)

Confidence	Page Nr.	Line transcription	Previ...
0.4983	3	des Mitherlehaer	
0.4707	1	dranz Miterlekner, dienskaachtsein	
0.2080	1	Moiveshekner Maurrts ded siusbesitzas	
0.1445	4	Jofuch M strlofet	
0.1429	5	Josiuch Mberlehent	
0.0574	1	z e Mensrgaaeeghdnetuhes. Msmernensgriduach Mti	
0.0530	4	t keeste seht Mnet lesnnehen Eni Me	

Preview



Close

Export document

Server export Client export

TEST_CITlab_Morchesne+_duplicated (123957)

Current document Current collection

Choose documents to export...

Finished server exports (not older than 2 weeks)

Choose export formats:

- Transkribus Document
- PDF
- TEI
- DOCX
- Simple TXT
- Tag Export (Excel)
- Table Export into Excel
- Export ALL formats
- Export Selected as ZIP

Export options:

Mets PDF TEI DOCX

- Export Page
- Export ALTO (Line Level)
- Export ALTO (Word Level)
- Export Image

Image type: Original

Filename pattern:

- pageNr + filename
- filename (warning: filenames must be unique for document)
- docId + pageNr + pageId

Pattern: \${filename}

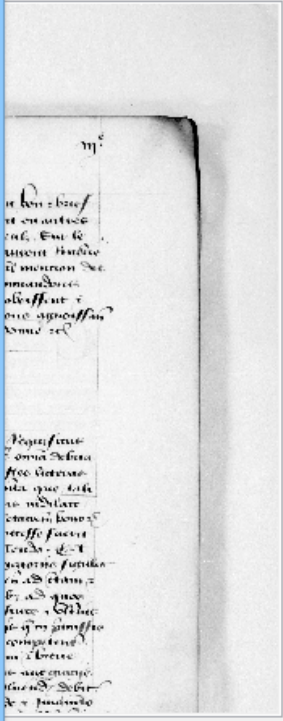
Placeholder: \${docId}, \${filename}, \${pageId}, \${pageNr}, \${filekey}

Version status: Latest version

- Word based
- Do blackening
- Create Title Page

Pages (5): 1-5 ... Current All

OK Cancel



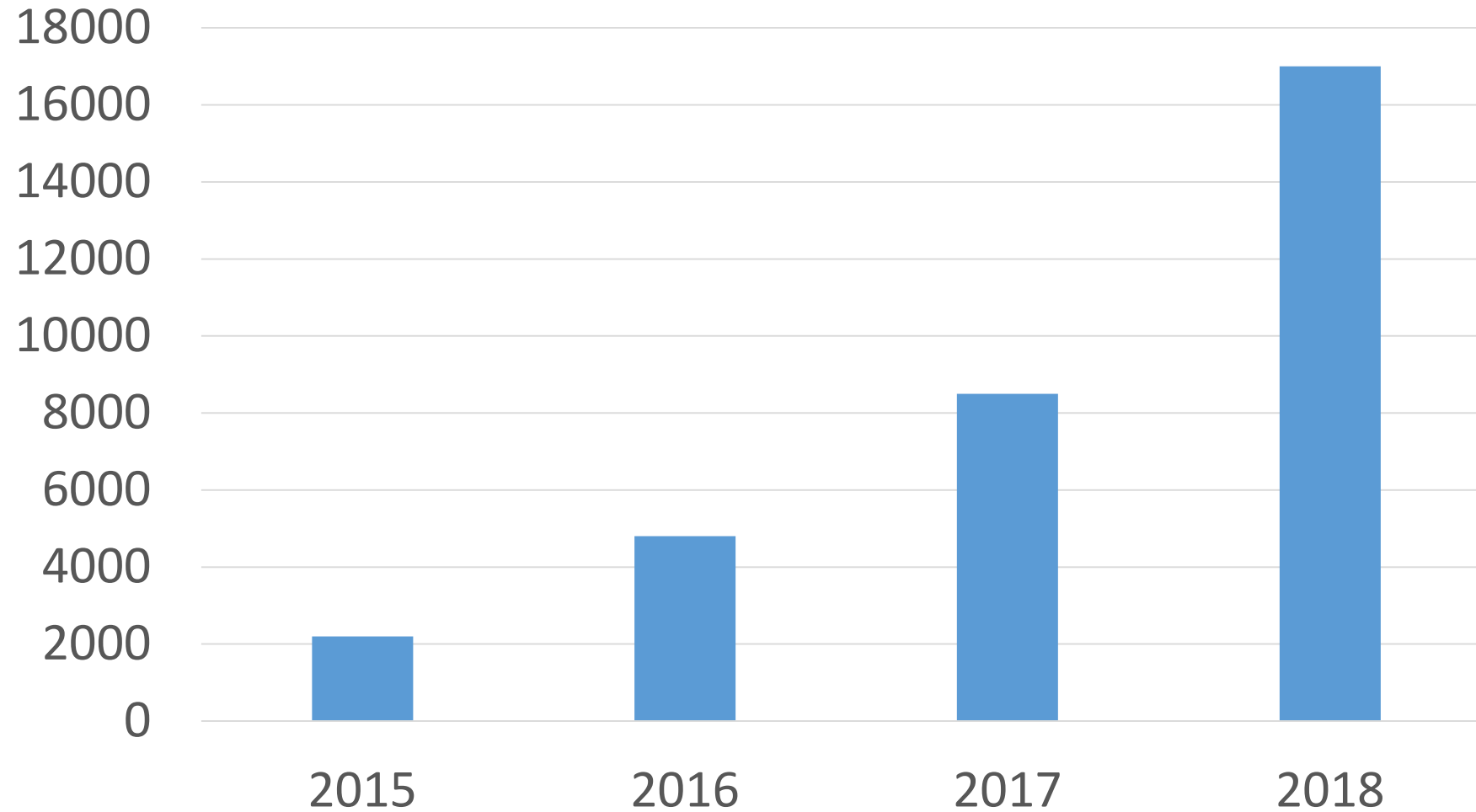
1-4 ipsi, tam conjunctim quam divisim in omnibus causis suis motis et
1-5 movendis contra quoscumque suos adversis coram quibuscumque iudicibus
1-6 regni nostri secularibus agendo et defendendo per procuratore seu actornatum

Transkribus - Plattform

Transkribus User Konferenzen – 2017 + 2018



Registrierte Benutzer in Transkribus



Woche vom 4. April bis 11. April 2019

- Images Uploaded by users: 98166
- New Users : 344
- Active Users / Unique Logins : 890
- Created Documents: 866
- Exported Documents: 230
- Layout Analysis Jobs: 1745
- HTR Jobs : 943

Trainingsdaten

- Jänner 2019
 - **228 HTR Modelle von Transkribus Benutzern trainiert worden**
- Trainingsdaten insg. in Transkribus (Februar 2019)
 - Seiten: **204.359**
 - Wörter: **21.200.035**
 - Ungefähr 120 Personenjahre an Arbeit
 - Monetärer Wert: ca. 2-3 Mill. EUR

Sharing is caring oder die Zukunft von Transkribus

Transkribus Zukunft

- Projekt endet mit 30. Juni 2019
- Allerdings ist die Nachfrage schon heute so groß, dass der Weiterbetrieb der Plattform bis Ende 2020 und darüber hinaus gesichert ist
 - EU Projekt NewsEye (2018-2021)
 - DFG Projekt Greifswald (2019-2020)
 - Projekt mit Nationalarchiv Finland (2019)
 - Projekt mit Nationalarchiv Niederlande (2019-2020)
 - Leuchtturmprojekt Tirol (2019-2020)
 - Projekt mit Trinity College Dublin (2019-2021)
 - Projekt mit Staatsarchiv Zürich (2019-2020)
 - Und weitere in Vorbereitung...

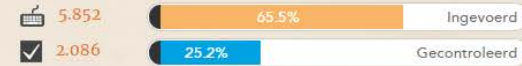


Crowd leert computer lezen

Leer de computer lezen en help mee aan het woord voor woord doorzoekbaar maken van de Amsterdamse Notariële Archieven.

Stand van zaken

📄 9.362 scans 👤 154 deelnemers



[Meedoen aan dit project](#)

Geschiedenis maken: oude handschriften & machine learning

In de donkere depots van het Stadsarchief Amsterdam ligt het archief van de Amsterdamse Notarissen 1578-1915: Miljoenen akten met daarin heel veel nog onbekende details over Amsterdammers en hun internationale netwerken. Over bekende en onbekende Amsterdammers en passanten, (internationale) handel, scheepvaart, slavernij, persoonlijk bezit en erfenissen, machtigingen en wisselprotecten, maar ook ooggetuigenverslagen van ontberingen op zee, ordinaire burenruzies of knokpartijen in herbergen en illegale speelhuizen.



Aan de slag

[Download de invoerinstrucies](#)

Projectdeelnemers



Software...

...will come and go, data will remain!

Warum nicht ein Geschäftsmodell entwickeln, in dem die Forschungsdaten – hier also das Wissen um historische Schriften und Dokumente – im Mittelpunkt stehen? Und das auf dem Gedanken der Zusammenarbeit und des Teilens beruht?

European Cooperative Society (SCE)

- Kooperative - Genossenschaft
 - Ermöglicht die Zusammenarbeit unabhängiger Institutionen um ein gemeinsames Ziel zu erreichen
 - Verteilte Eigentümerschaft und das Teilen von Daten stehen im Mittelpunkt
- Wichtige Eigenschaften einer SCE
 - Offen für neue Mitglieder, geringe Hürde: 1000 EUR Anteil als Minimum
 - Demokratische Konstitution: Verwaltungsrat - Generalversammlung
 - Direkter Vorteil für Mitglieder als Ziel – kein Shareholder Value
 - Kunden werden Eigentümer, Eigentümer werden Kunden
 - Subskriptionsgebühren und servicebasierte Gebühren

Derzeitiger Stand

- Statuten
 - Weitgehend fertig gestellt
 - Verwaltungsrat wird in den nächsten Wochen geformt
 - Gründungsakt soll vor dem Sommer stattfinden
- Gründungsmitglieder
 - Universität Innsbruck, Universität Greifswald, Technische Universität Valencia, National Archiv Finland, British Library, Universitätsbibliothek Belgrade, Diözesan Archiv Passau, Universität Rostock, ZAMG Wien, Geneanet Frankreich, etc..
- Jede Institution, die mit Transkribus arbeiten möchte, ist herzlich eingeladen mitzumachen!

Vielen Dank für die Aufmerksamkeit

Weitere Informationen

<https://read.transkribus.eu/>

<https://transkribus.eu/>

<https://read.transkribus.eu/coop/>

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943.



<http://scantent.eu/>

