



**Kanton Zürich
Direktion der Justiz und des Innern
Staatsarchiv**

StAZH

Recognizing Handwriting

Large and Small Scale

Tobias Hodel

READ





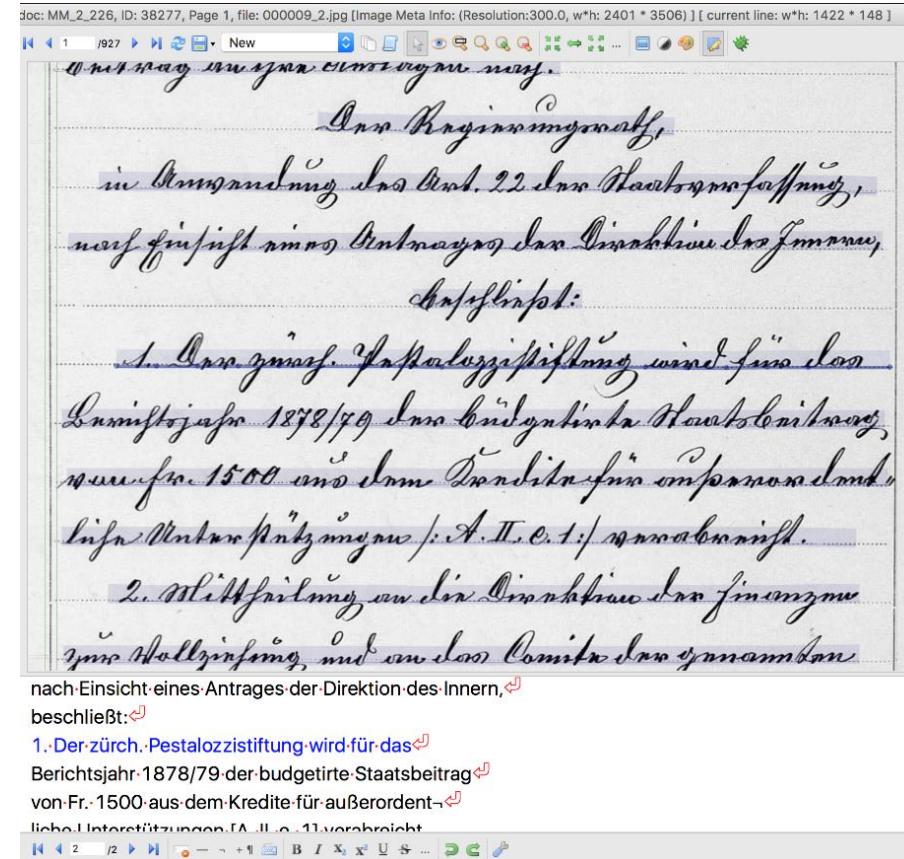
Regierungsratsbeschlüsse GT of 150'000 pages

- German Kurrent
- Spanning 19th century (1803-1883)

Preparation:

- t2i (by URO)
- Success in 90% of lines

→ Ready for training of HTR models



Bundesratsprotokolle READ/StAZH – Swiss National Archives

Project to recognize 150'000 pages

- Quality of LA and HTR sufficient
- Based on GT from state archives

Goals:

- Searchability (KWS/Indexing)
- Text Extraction
- (potentially:) Document Understanding

→ HTR+ as game changer



READ

<p>Präsid. Verfügungen v. 14. Januar 1894.</p> <p><u>Politisches.</u> Antrags n. 30. vor. Mitt. 1. Johann Martin <u>Ankele</u> von Gomadingen, Württemberg, in Singen (Hohentwiel) Schaffhausen; 2. Johann <u>Bär</u> aus Frankreich, in Runggarten, Augsburg; 3. Josef Charles <u>Baselys</u> aus Frankreich, in Genf; 4. Francesco Dominico <u>Faglardi</u> von Chiuffi, Itali., in Genf; 5. Johann Friedrich <u>Leitz</u> von Nürnberg, in Leitz; 6. Jacques <u>Nellici</u> von Castello d'Amone, Itali., in Düsseldorf; 7. François Joseph <u>Rambozson</u> aus Frankreich, in Charente, erhält die Bewilligung zum freien und öffentlichen Antritt und Gemeindeeingangsstelle. On die Petitionen. Ad 1) An Schaffhausen z.R. Pol. Antrag auf geleit. Registrierung z.R.</p>	1
<p><u>Politisches.</u> Antrags n. 29. vor. Mitt. Das Obertribunal der 'Dietia de la Republica Mayor de centro America' teilt mit, daß das Besondere infolge der Vereinigung des drei zentralamerikanischen Republikanischen Bundesstaates, Nicaragua und Salvador befohlen habe, die den bei den Gross-amerikanischen Konföderationen bestehenden Gewerbeaufgaben, welche die beiden Staaten die spätere Etagen Lebet, bis zur Amtzeit von Honduras und Nicaragua in Basel, Benjamin Raat, bis zur Amtzeit, und Nathan Karr, bis zur Amtzeit von Salvador in Genf, und der Lizenzen der ausländischer Anwälte in den Büros geöffneten. On Baselstadt und Genf z.R. On die Dietia der Republica Mayor in San Salvador z.R. Jede Landeshälfte pol. Antrag auf geleit. Registrierung z.R.</p>	2
<p><u>Politisches.</u> Antrags n. 30. vor. Mitt. Christian <u>Doege</u> von Altenbergen, Preußen, in Lecke, am 1. November 1893 für Bewilligung zw. vor - der Doege Aufstellung der Bürgerschaftsabstimmung.</p>	3

HTR+

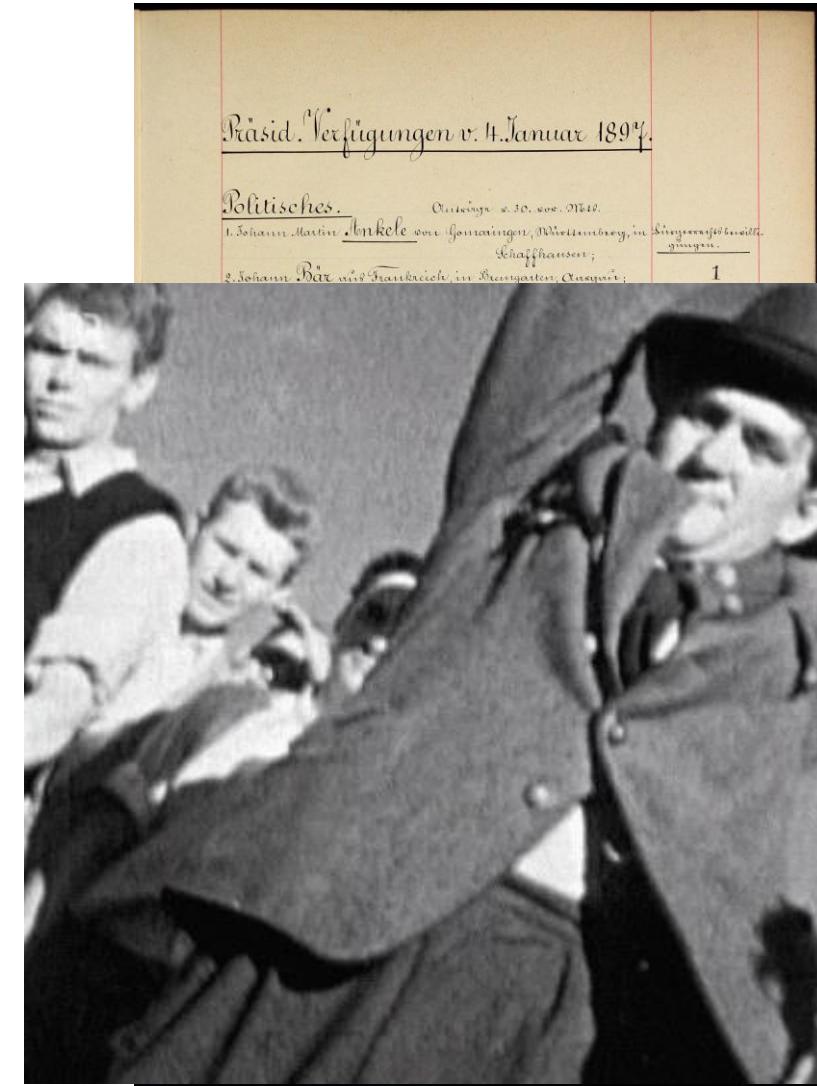
- Model based on StAZH GT (800 pages, spread over century):
CER around 7%
- Applied on Bundesratsprotokolle:
CER slightly higher, around 9%
- 6-10% lower compared to currently implemented HTR

HTR+

... Significantly better → quality

.... Works on similar handwriting not being part of training set → adaptability

even beyond ...



allein jenes Mikroplast sei selbst Lungenfunktionen überzählig,
denn das Präsidium der Wissenschaften sei, in Größen
der Organisations in hingeworfen Aufgaben ausserordentlich
groß. Das Landesamt bildet in eigentlicher Form
das Präsidentenamt des Landes und ist der Landesbehörde,
seiner Aufgabe mithin mehr in der Pflicht als im Wege

allein jedes Mitglied sei selbst davon seinen überzeugt, ↗
das des Präsidiums der Micha überhoben sei, die Große ↗
der Gegenwart in langerer Ansprache auseinander zu ↗
hetzen. der Bundesrat bilde in eigentlichen Sinne ↗
des Wortes den praktischen Ausschuß der Bunderbehörde ↗
dessen Aufgabe mithin mehr in der hat als im Worte ↗
bestehen. Das hroisidun heiße daher die Bundesräthe ein ↗
nach willkommen, indem es den Wunsch beifugge, das stets ↗
ein freundliches und loyales Verhältniß vorherrschen morgen. ↗
Zwar, wenn von einträchtigem Zusammenwirken in ↗
sprachen wurde, so sei damit nicht eine durchaus im

HTR+

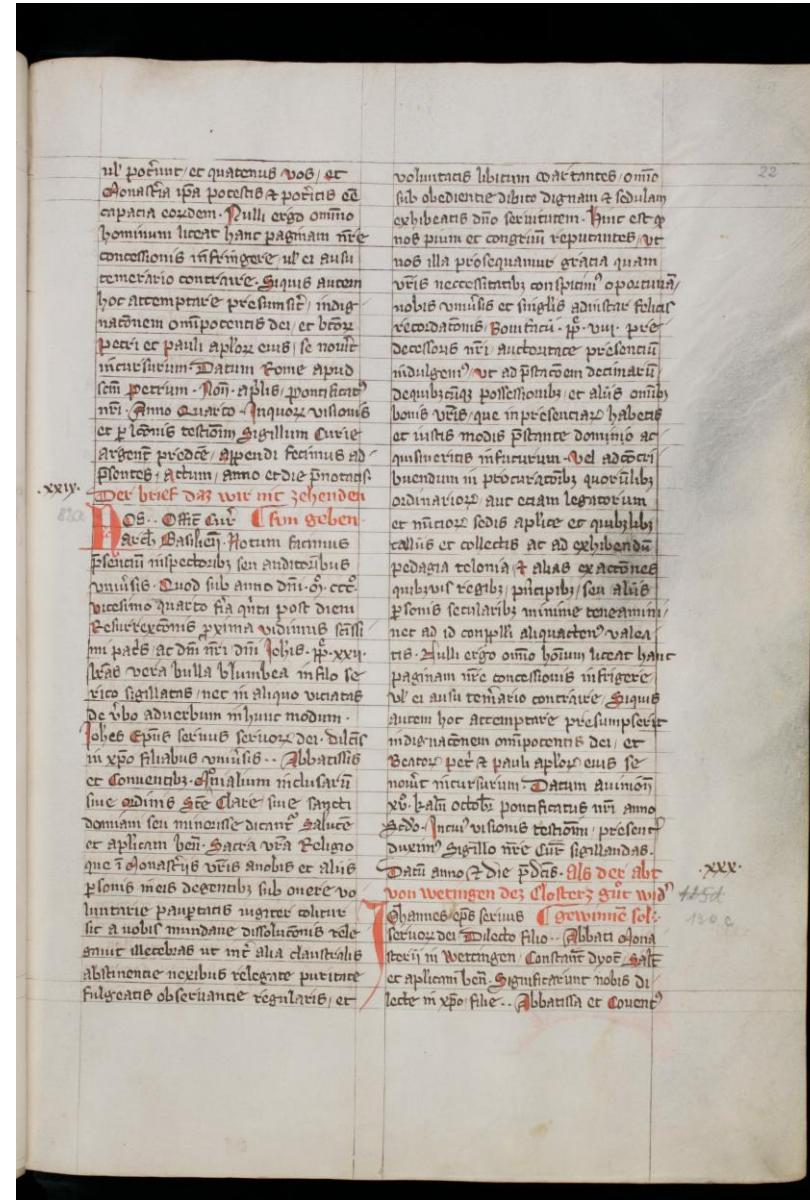
- Model based on Passau GT (1'120 pages, spread over 19th century): delivers similar results (CER about 1-2% worse)!

Generic HTR models can be gained, using HTR+ for

- ... German Kurrent (19th century)
- ... Gothic Handwriting (13th to 16th century)

Potentially:

- ... Charters (12th to 17th century)
- ... everything there is enough GT for!



Handwritten Finding Aids StAZH

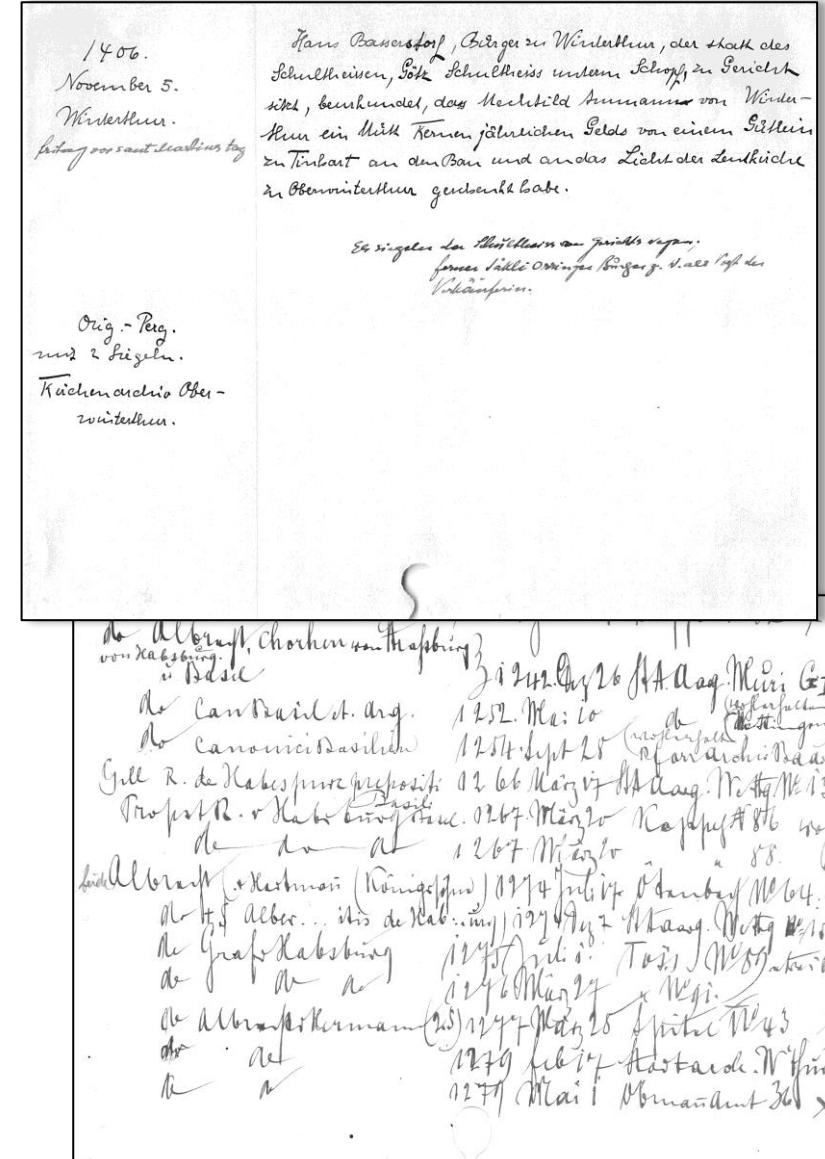
Recognition of Handwritten Finding Aids

- Regesta of Charters
- Index of Letters (Sender, Receiver, Date)
- Old Archival Finding Aids (elaborated in 19th century)

Ingested in Transkribus and Searchable via KWS
(for internal use only).

Currently carried out for State Archives of Berne:

- KWS of birth, death and marriage registers



READ

vof dom. zu jumij ist. thoma
guimtr und sime. margret mißmi
gehofft : . erist. Els bet. 3ügern.
Mawti. löffer. Catrin. grima
ma. Appolonia. Sigib.

✉

e. Alteguns, Jannfaer. kguser ✉

gehaufeeis. 2e. bet. beugern ✉

gde ✉

gele ✉

✉

✉

arsten. latscher, ateazimme. ✉

1. Ds. ✉

tich ✉

. 9cf. ✉

Transkribus: transkribus.eu / email@transkribus.eu / transkribus.eu/wiki/

READ: read.transkribus.eu (auch für News)

Staatsarchiv Zürich: tobias.hodel@ji.zh.ch // Twitter: @thist_uzh



Ratsmanuale (Council Minutes) 1'000 Volumes at StAZH (1484-1798)

Starting second half of 2019 through 2022

Project to recognize 180'000 pages

- Training of LA necessary
- GT for different scripts needs to be produced
- Evaluation of DU (identification of pages, titles, and marginalia)

Goals:

- Searchability (KWS/Indexing)
- Iterative Correction of Text by Crowdsourcing and better HTR models



READ

