

# “Ein Freinttlichen Trunck”: The Zechzettel Dataset for Handwritten Text Recognition and Information Retrieval

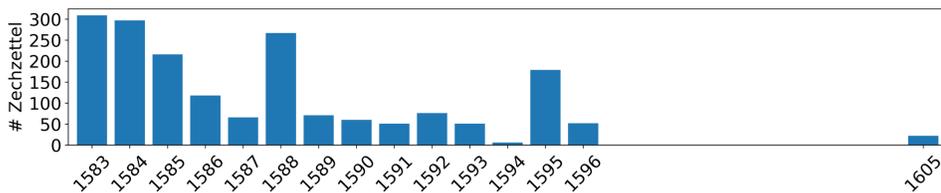
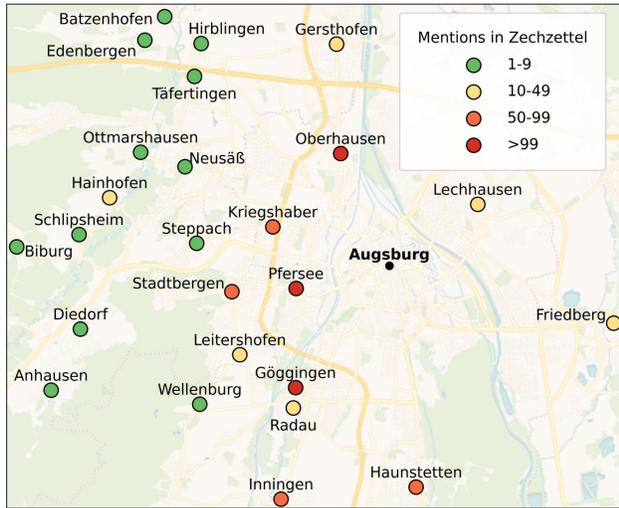
Fabio Mariani<sup>1</sup>, Helmut Graser<sup>1</sup>, B. Ann Tlusty<sup>2</sup>,  
Regina Dauser<sup>1</sup>, Annemarie Friedrich<sup>1</sup>

<sup>1</sup>University of Augsburg, Germany; <sup>2</sup>Bucknell University, Pennsylvania, US

## Historical Context and Data Sources

The Zechzettel dataset is based on a historical collection of around 1,830 handwritten petitions produced in Augsburg (Germany) between 1583 and 1605, and preserved at the Stadtarchiv Augsburg.

These petitions requested permission for drinking tax-free outside the city limits, a practice regulated by local authorities.



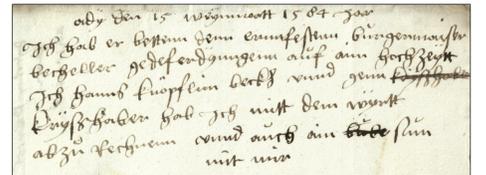
## Metadata and Annotations

### Handwriting Quality



“Gepflegte kleine Notarshand”  
(David Edelstein, 16.10.1584)

Professional Notarial  
Small



“Bemühte, unausgeglichene, große Hand; wohl eigenhändig”  
(Hanns Knöpflin, 15.10.1584)

Non-Professional Irregular  
Large

### Named Entities

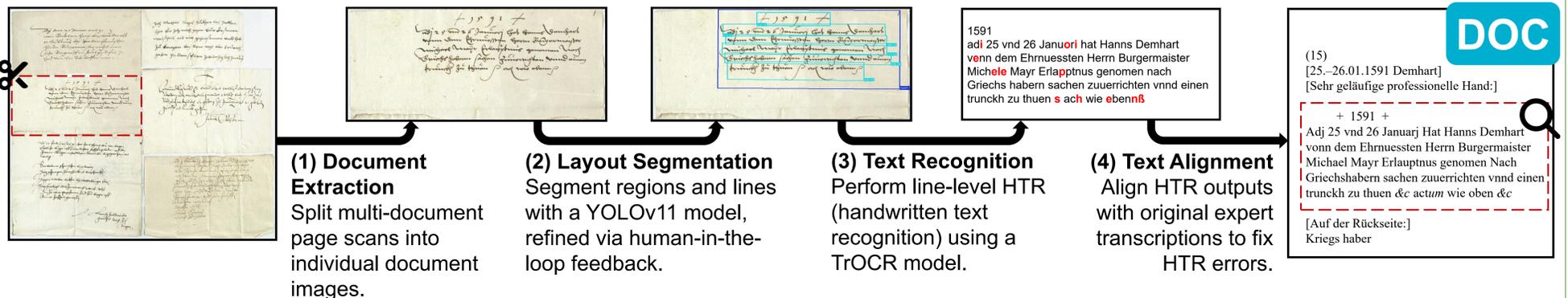
+ 1591 +  
Adj 25 vnd 26 Januarj Hat Hanns Demhart  
vonn dem Ehrnuessten Herrn Burgermaister  
Michael Mayr Erlauptnus genomen Nach  
Griechshabern sachen zuerrichten vnd einen  
trunckh zu thun &c actum wie oben &c

### Historical Variants:

Griechshabern / Kriegshaber /  
Kriegshabern / Grieshabern /  
Kriehshabern / ...

Modern Name: Kriegshaber

## From Sources to Dataset



## Background Corpus for Baseline HTR Models

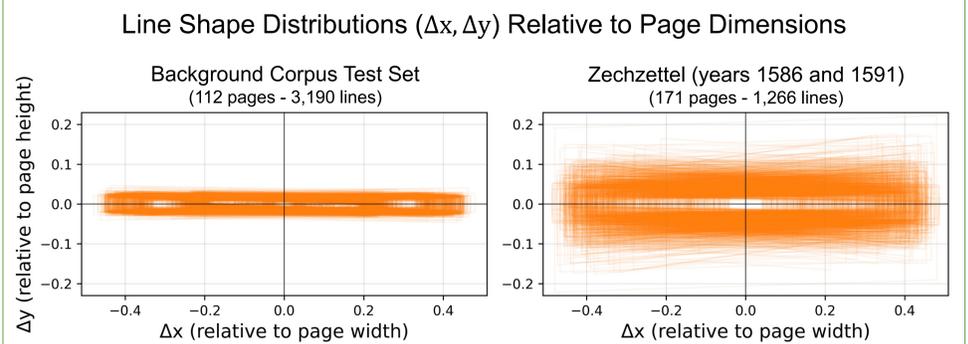
Layout Segmentation: YOLOv11-Large (fine-tuned)

Text Recognition: microsoft/trocr-large-handwritten (fine-tuned)

Collection	Century	Pages ( 1 117)			Lines (30 282)		
		Train	Validation	Test	Train	Validation	Test
Ratsprotokolle <sup>1</sup>	17th	360	45	45	8 590	930	1 024
Reiseberichte <sup>2</sup>	17th	52	7	7	1 230	217	194
Dresdner Hofdiarium <sup>3</sup>	17th	34	4	4	742	91	82
Konzilsprotokolle <sup>4</sup>	18th	264	33	33	6 925	881	861
Senatsprotokolle <sup>5</sup>	19th	183	23	23	6 653	833	1 029
<b>TOTAL</b>	—	<b>893</b>	<b>112</b>	<b>112</b>	<b>24 140</b>	<b>2 952</b>	<b>3 190</b>

[1] Toselli, A. H., Romero, V., Villegas, M., Vidal, E., & Sánchez, J. A. (2018). *HTR dataset ICFHR 2016*. Zenodo. <https://doi.org/10.5281/zenodo.1164045>  
 [2] Mariani, F. (forthcoming). *Reiseberichte HTR dataset* (derived from: Wenzel, M., Jagodzinski, S., & Timann, U., *Philipp Hainhofer: Reiseberichte und Sammlungsbeschreibungen* (München 1603; Stuttgart 1621), Wolfenbüttel Digital Editions, No. 4)  
 [3] Beckert, S. (2024–2025). *Ground truth sets for handwritten text recognition (HTR/OCR): Dresdner Hofdiarium (17th-century Kurrent manuscripts)*. Zenodo. <https://doi.org/10.5281/zenodo.14356190>; <https://doi.org/10.5281/zenodo.15303243>; <https://doi.org/10.5281/zenodo.15303398>  
 [4] Grüning, T., Leifert, G., Michael, J., Strauß, T., Weidemann, M., & Labahn, R. (2016). *Read\_dataset\_german\_konzilsprotokolle*. Zenodo. <https://doi.org/10.5281/zenodo.215383>  
 [5] Tübingen University Library. (2023). *Senatsprotokolle*. GitHub. <https://github.com/ubtue/Ground-Truth/tree/main/Senatsprotokolle>

## Domain Shift: Zechzettel vs Background Corpus



## Preliminary Layout Segmentation Results

Model BG: YOLOv11-Large fine-tuned on Background Corpus (BG)  
Model BG+Z: Model BG fine-tuned on 120 Zechzettel (year 1586)

Model	Test Set	Precision	Recall	mAP@50
BG	BG Pages Test Set	0.96	0.91	0.96
BG	51 Zechzettel (year 1591)	0.79	0.85	0.85
BG+Z	51 Zechzettel (year 1591)	0.97	0.95	0.97



Contact:  
Fabio Mariani  
Lehrstuhl für Computerlinguistik, Universität Augsburg  
fabio.mariani@uni-a.de

Project Repository  
[github.com/zechzettel](https://github.com/zechzettel)

