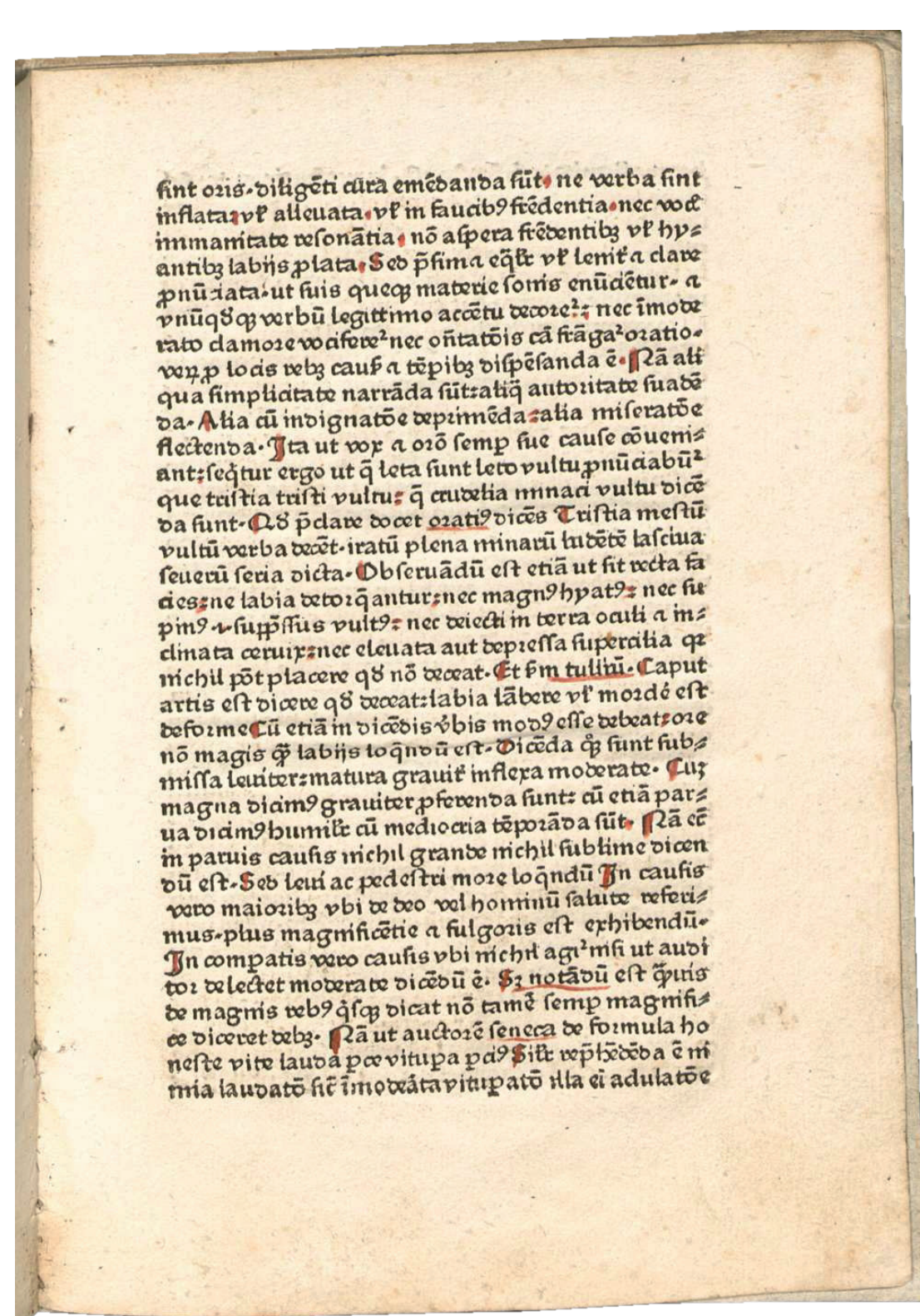


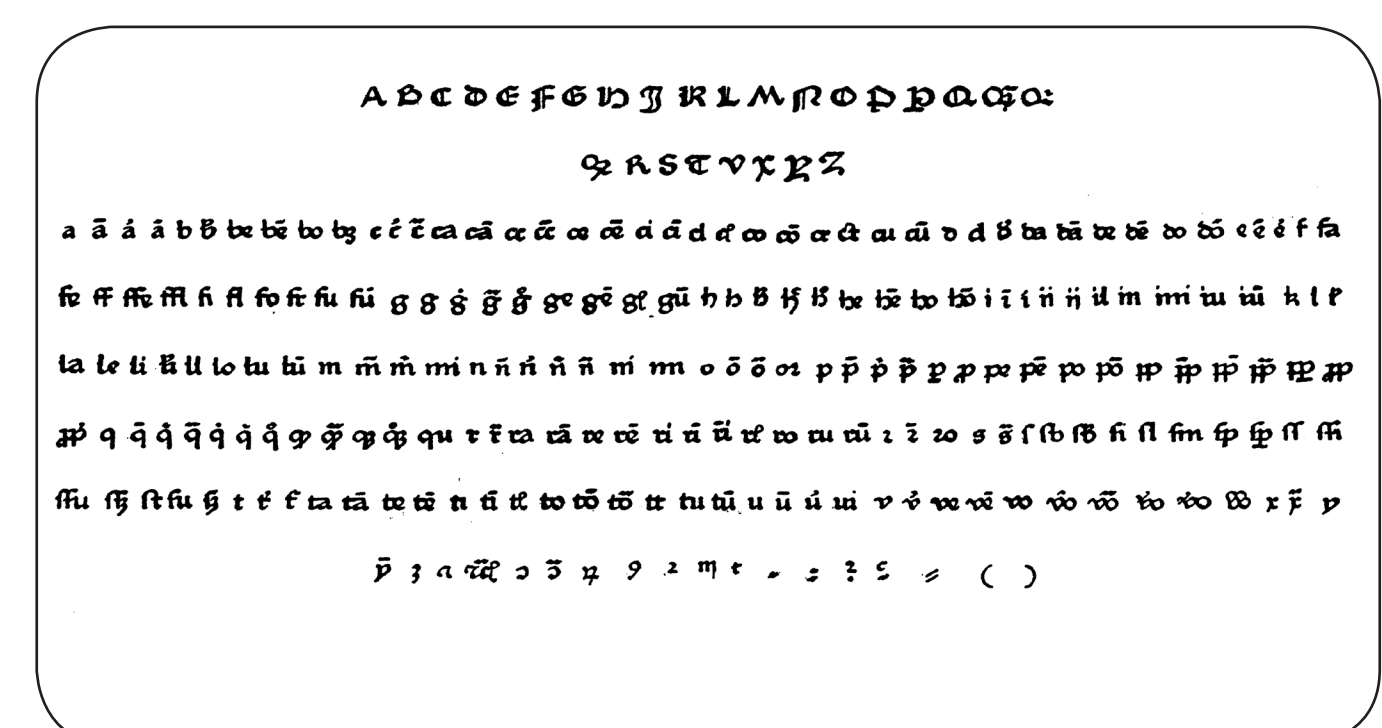
Automatische Typenerkennung in historischen Drucken

Anwendungsfall

- Frühdrucke werden u. a. über die verwendete Schrift datiert.
 - Das dafür übliche Verfahren wurde Anfang des 20. Jahrhunderts entwickelt.
 - Es basiert auf der Identifikation der Form von >M< bzw. >Qu< einer Schrift sowie der Schriftgröße über 20 Zeilen.
 - Mithilfe dieser beiden Faktoren kann die Schrift näherungsweise eingegrenzt werden. Zwischen den verbleibenden Möglichkeiten wird entschieden, indem die Zeichen des vorliegenden Drucks mit Alphabettafeln aus dem Typenrepertorium abgeglichen werden.
- Dieses Verfahren ist langwierig und nur durch Experten zu verwenden.



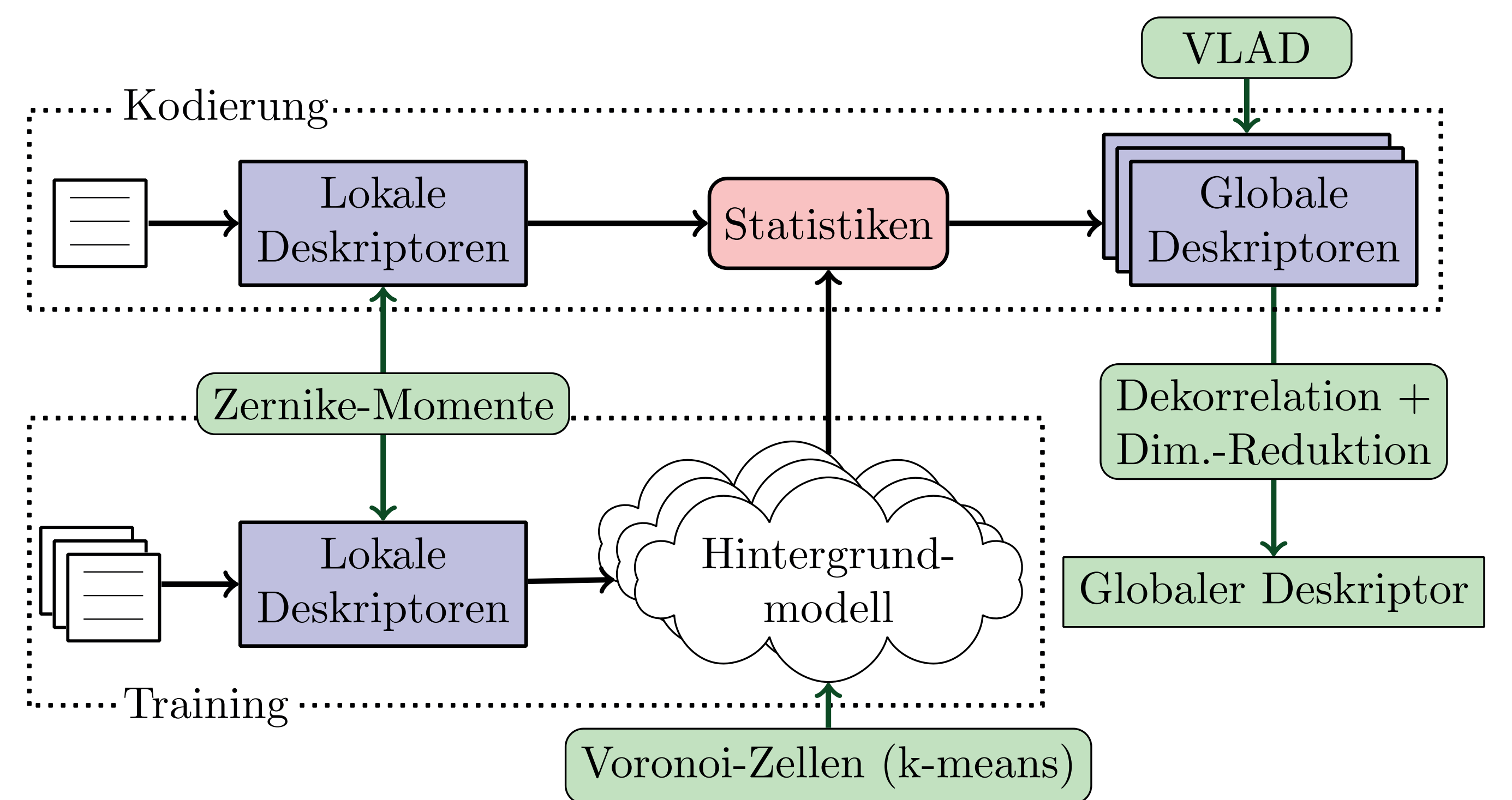
M	M	M	M	M	M	M	M	M	M
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100



Ulrich Zell
Köln
1480–1491

Type 3:83G

Methodik



V. Christlein, D. Bernecker and E. Angelopoulos, »Writer Identification Using VLAD Encoded Contour-Zernike Moments,« Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, Nancy, 2015, pp. 906–910.

Evaluation

- Erster Test an 100 Bildern
- Berechne für jedes Bild einen globalen Deskriptor
- Teste jeden Deskriptor mit den 99 anderen und prüfe ob gesuchte Type unter den ähnlichsten Dokumenten in der Top X Liste (leave-one-document-out cross validation)

Top 1	44 %
Top 5	62 %
Top 10	72 %
mAP	52 %

Herausforderungen

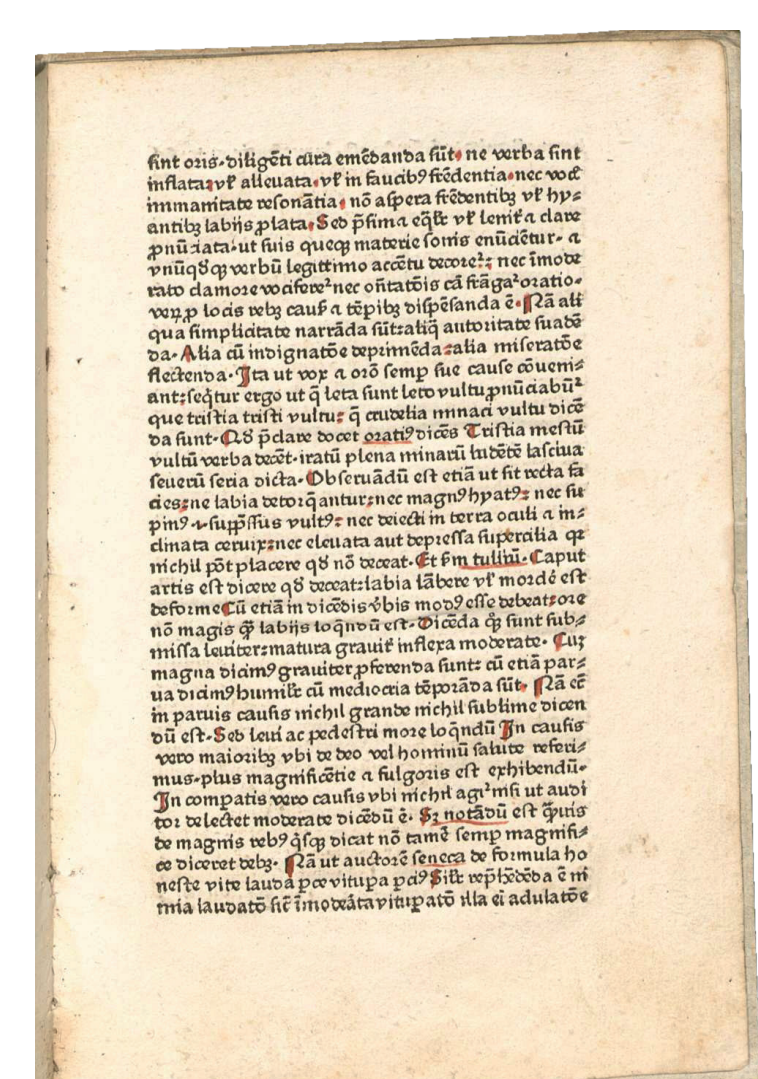
- Unterscheidung von Schrift und Bedruckstoff
- Stark schwankende Qualität von Digitalisaten
- Differenzierung bei der Verwendung von mehreren Drucktypen
- Typen- und Datenmenge: Mehr als 6.000 Drucktypen und über 500.000 Seiten Digitalisate.

Zusammenfassung

- Erste erfolgreiche Anwendung einer automatischen Typenidentifizierung
- Durch mehr Trainingsmaterial werden noch bessere Ergebnisse erwartet
- Potenzial für die Verwendung der Typenerkennung auch für die Zeit nach 1500

Zielsetzung

- Ergänzung der Methode durch einen mustererkennungsbasierten Ansatz
 - Die bestehende Zuordnung von ca. 15.000 digitalisierten Inkunabelaufgaben zu den jeweiligen Typen im Typenrepertorium bildet einen herausragenden Bestand an *Ground-Truth-Daten*.
 - Diese sind Grundlage für Entwicklung und Training eines Algorithmus, der mit hoher Zuverlässigkeit Schriften vollautomatisch erkennt.
- Beschleunigung der Typenbestimmung für Spezialisten und Ermöglichung des Verfahrens für viele Nicht-Inkunabelspezialisten.



direkte Erkennung →

Type 3:83G

Kontakt / Projektpartner

JGU Mainz
Institut für Buchwissenschaft
Dr. Nikolaus Weichselbaumer
weichsel@uni-mainz.de

FAU Erlangen-Nürnberg
Lehrstuhl für Informatik 5
Mustererkennung
Dipl.-Inf. Vincent Christlein
vincent.christlein@fau.de