

Updates zum Projekt „Robuste und performante Verfahren für die Layoutanalyse in OCR-D“

Christian Reul und Robert Sachunsky

Forschungsstelle DACHS, Universität Würzburg
SLUB Dresden

19.11.2025



Agenda

1. Projektübersicht

2. LAREX (Teaser)

3. Trainingsdaten

4. Trainings-Testereien

5. (Meine) Fragen

Übersicht

- Förderung durch DFG
- Förderlinie e-Research-Technologien
- Laufzeit 2023 bis 2026 (März)
- Partner
 - Staatsbibliothek zu Berlin
 - SLUB Dresden
 - ZPD Würzburg

Ziele

- (Weiter)Entwicklung von Tools ...
 - zur automatischen Segmentierung: Eynollah und Detectron-basiert
 - zur Visualisierung und händischen Auszeichnung/Korrektur: **LAREX**
- **Erstellung und Konsolidierung von Trainingsdaten**
- Entwicklung und Implementierung neuer Evaluationsmethoden
- **Herumexperimentieren mit Trainingsworkflows etc.**

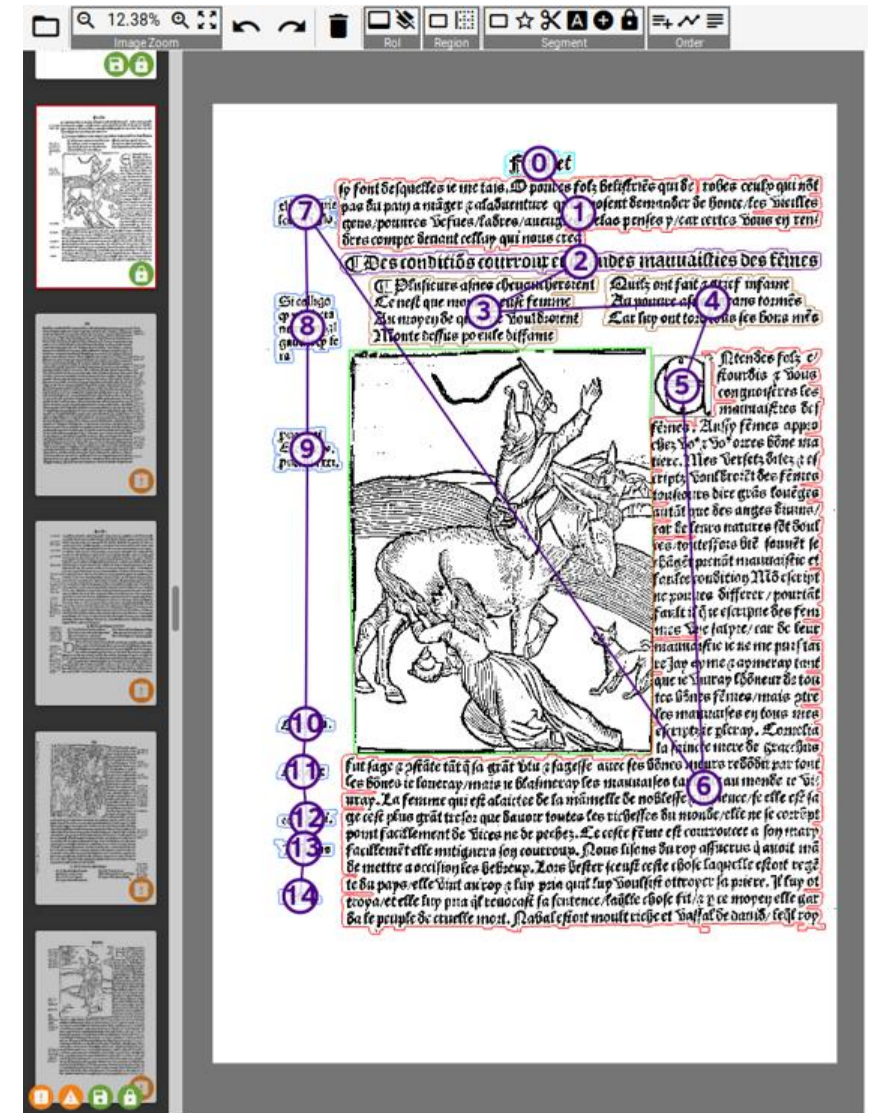
Agenda

1. Projektübersicht
- 2. LAREX (Teaser)**
3. Trainingsdaten
4. Trainings-Testereien
5. (Meine) Fragen

Übersicht

Umfangreiches Tool zur Anzeige und Korrektur von quasi allen (Zwischen)Ergebnissen eines ATR-Workflows, u. a.

- Regionen- und Zeilenkoordinaten
- Semantische Typisierung von Regionen
- Reading Order
- Text



Alte Schwächen

- Eigentlicher Plan im Projekt: Punktuelle Ergänzung von Funktionalitäten, u. a.
 - Komplexere Reading Order
 - Metadaten
- Aber: LAREX lange historisch gewachsen
 - Eher für die Einzelnutzung konzipiert (keine Nutzerverwaltung etc.)
 - Selbst kleinere Änderungen oft aufwendig
 - Tech-Stack in die Jahre gekommen
 - ...
- Also: Komplette Neuimplementierung

Neue Stärken

- Moderner Tech-Stack, deutlich höhere Flexibilität und Modularität
 - Backend: Spring Boot 3 + Spring/Java Ecosystem (Spring Security, Spring Data JPA, etc.)
 - Auth: Keycloak
 - Frontend: Nuxt 4 + nuxt-auth-utils (secure OAuth2 login via BFF pattern) + WebGL Editor
- Features
 - Umfangreiche Nutzer- und Datenverwaltung
 - Multiview
 - Flexibler Input (multipler) Daten (.png / .nrm.png / .bin.png / ... ; .xml / .workflow1.xml / ...)
 - Erweiterte Reading Order Funktionalität (insb. ordered/unordered groups)
 - Umfangreicher Metadaten-Editor
 - ...

Agenda

1. Projektübersicht
2. LAREX (Teaser)
- 3. Trainingsdaten**
4. Trainings-Testereien
5. (Meine) Fragen

Übersicht

- Viele unterschiedliche Quellen
- Fokus auf „OCR-D-relevante“ Daten
- (Ursprüngliche) Grundidee: Training in Pipelines
 - Frühe Schritte in der Pipeline
 - Robustheit
 - Qualitativ schlechter / allgemeinere Daten / ...
 - Spätere Schritte in der Pipeline
 - Finetuning
 - Qualitativ bessere / passendere Daten / ...
- (Ursprüngliche) Hoffnung: Auch eher fragwürdige Daten haben relevanten Nutzen

Segmentierungsrichtlinien

- Generelle Vorgehensweise
 - So ausdifferenziert wie nötig
 - So allgemein wie möglich
- Überarbeitung der OCR-D GT-Guidelines:
https://ocr-d.de/de/gt-guidelines/trans/struktur_gt.html
- „Finetuning“ anhand von Erkenntnissen aus anstehenden Evaluationen
- Aktueller Eindruck
 - Stark abhängig vom individuellen Anwendungsfall
 - One-fits-all Lösung unwahrscheinlich
 - Leichte Anpassbarkeit sicherstellen (Skripte, ...)

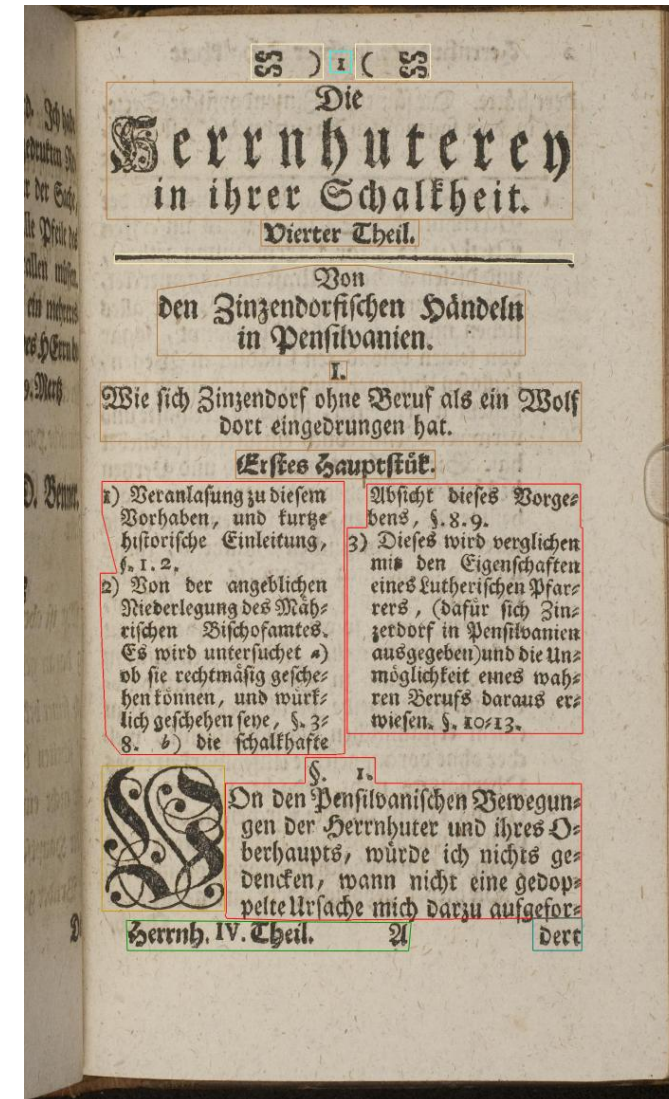
Quellen

- OCR-D Referenzdaten
- DTA
- PRImA
- Schulbücher
- (Riesige synthetische Datensets)

OCR-D Referenzdaten

- <https://ocr-d.de/de/daten>
- Gezielte Auswahl aus
 - DFG-Projekt „Deutsches Textarchiv“
 - Digitalisierten Sammlungen der Staatsbibliothek zu Berlin
 - Wolfenbütteler Digitalen Bibliothek der Herzog August Bibliothek
- Enthält neben Layout- auch Text-GT

OCR-D Referenzdaten



DTA

- https://github.com/OCR-D/gt_structure_all
- Riesige Datenmenge: 25.441 Seiten
- Nutzen und Nutzbarkeit der Daten schwankt allerdings stark
 - Sehr hoher Anteil von Standardseiten
 - Stark schwankende Auszeichnungsqualität
- Auswahl für genauere Prüfung/Korrektur durch Heuristiken, u. a. → ca. 3.100 Seiten
 - Polygone
 - Regionentypen
- (Semi)Automatische Anreicherung mit Zeilen- und Baseline-Koordinaten
 - Basierend auf https://github.com/bertsy/ocrd_cis
 - (Semi)Automatische Qualitätsanalyse angedacht

DTA – Aussortiert

6

Erstes Buch. Geschichte der Teutschen

Sohn in den Eilfischen Feldern vorher verkündiget.⁶ Auf einer Münze des Corneli-
schen Hauses finden wir diesen Marcellum abgebildet, wie er die Waffen, so er
dem erschlagenen Virdomarus abgenommen, in dem Tempel des Iouis Feretrii auf-
stellt.⁷ Nur ist es zweifelhaft, ob diejenigen, so die Fastos Capitolinos abge-
schrieben, recht gelesen, indem unter allen Historicis keiner die Teutschen neu-
net, und der Name Germani zu der Zeit wohl noch nicht bekannt gewesen. Viel-
leicht haben sie Germanis, anstatt Cenomanis gelesen: diese waren ebenfalls ein
Gallisches Volk, so nächst den Insulern das Land, wo nachmals Brescia, Crema-
na und Piacenza angelegt worden, bewohnet, und so wohl an diesem, als nach-
maligen Kriegen gegen die Römer Theil gehabt. Wie wir denn in eben diesen Fa-
stis finden, daß nicht lange nach Marcello, Cethegus über beyde zugleich
triumphiret.⁸

V. Deutlichere Nachrichten finden wir von den Colonien, so die
Teutschen an jene Seit des Rheins nach Gallien geschickt. IVLIVS CAESAR¹
berichtet, daß die Belgen, so den dritten Theil von Gallien innen hatten, und zu sei-
ner Zeit die tapfersten unter den Galliern waren, alle Teutscher Abkunft gewesen.
Die Trevirer und Nervier insonderheit wußten sich viel damit, daß sie von Teut-
schen abstammten. Die Belgen haben hernach die Küsten von Britannien be-
völkert² und beständig mit den Germanis zu schaffen gehabt.³

VI. Selbst der Name Germani ist an der andern Seite des Rheins zu-
erst

Erstrecken der
Teutschen, so
sich in Gallien
niederlagerten

6. VIRGILIUS AEN. VI. 860.
Aspic ut insignis spolia Marcellus opimis
Incedit, nullusque virus supereminet omnes.
Hic rem Romanam magnos turbante tumultus
Sistet equos, sternet Poenos, Gallumque rebellem
Terrique arma patri, suspendens capta Quirinus.
Confer, PERZONIVM in minimaducti Hist. c. 7

7. V. PATINI familiae Romanae in numis
p. 87. num. 3. 4.

8. Ad an. V. C. DEVI. C. CORNELIVS
L. P. M. C. THEGVS COL. DE INVERIVS ET CENO.

9. V. I. CAESAR. L. II. c. 4. Quom ab his quaer-
veret, quid in bello possent. . . sic reperiebat, plera-
que Belgas esse ortos a Germanis, Rhenumque anti-
quius transiitibus, propter loci fertilitatem hic con-
fessit. Gallosque, quos loca incolerent, expulsi-
se: solosque esse, qui patrum nostrorum memoria
omni Gallia uocata, Testunet, Cimbrumque intra
fines suos ingressi praesiderent, quae ex re ferri, ut
earum verum memoria magnam sibi aueritatem,
magnaque spiritus in re militari sumerent.

2. IDEM L. V. c. 12. Britanniae pars interior ab
his incolitur, quos mater in insula ipsa memoria pre-
ditum dicitur. Maritima pars ab his, qui, praedat
ac belli inferendi causa ex Belgis transierunt, qui
antei fere in nominibus ciuitatem appellantur, qui-
bus orti ex ciuitatibus ex peregrinis, et bello ille-
gitis his romanis fuerunt, atque agris colere coeperunt.

TACITVS VII. Agricola c. 11. Ratiuae Cal-
doniam inhabitantium Comas, magni artus, Gera-
manicus originem asserunt.

3. CAESAR. L. I. c. 4. Horum omnium
fortissimi sunt Belgae, propterea quod a cultu atque
humanitate prouinciae longissime abstant, minime
que ad eos mercatores seque committunt, atque ea,
quae ad efficiendum animi pertinent, important.
Proximique sunt Germanis, qui trans Rhenum ince-
dunt, quibus cum continentem bellum gerunt.

§. VI. 1. IDEM L. II. c. 4. Condrasus, Ebor-
aci, Caracasi, Pannoni, qui uno nomine Ger-
mani appellantur, & L. VI. c. 31. Segni, Condrusi-
que, ex gente & numero Germanorum, qui sunt in-
ter Eboracum, Treuironem, legatus ad Caesarem
misit, et ait, ne se in bellum numero deceret,
nec omnium Germanorum, qui essent citra Rhe-
num, causam esse unam indiceret.

2. TACITVS de M. G. c. 2. Ceterum Ger-
manis incolentibus rectus & iugis solitudo: quon-
iam, qui primi Rhenum transgressi, Gallos expulerunt
ac nunc Tungri, tunc Germani uocati sunt. Ita
nationis nomen, non gentis euasisse paulatim, ut
vixit primus a uillare ut metum, mox a se ipso,
inuenito nomine, Germani uocarentur. Dicit Cicero,
wie sie an sich bündelt, hat den Gelehrten, die den
Ursprung des Namens Germani erörtern wollen, in
vielen, zum Theil sehr gesungenen Erklärungen be-
halten müssen, denen sie gleichwohl keine alte & ein-
zigste ist. Es erhellet aber doch zum wenigsten so viel
daraus

6. Virgilius Aen. VI. 860.
Aspic ut insignis spolia Marcellus opimis
Incedit, nullusque virus supereminet omnes.
Hic rem Romanam magnos turbante tumultus
Sistet equos, sternet Poenos, Gallumque rebellem
Terrique arma patri, suspendens capta Quirinus.
Confer, PERZONIVM in minimaducti Hist. c. 7

7. V. PATINI familiae Romanae in numis
p. 87. num. 3. 4.

8. Ad an. V. C. DEVI. C. CORNELIVS
L. P. M. C. THEGVS COL. DE INVERIVS ET CENO.

9. V. I. CAESAR. L. II. c. 4. Quom ab his quaer-
veret, quid in bello possent. . . sic reperiebat, plera-
que Belgas esse ortos a Germanis, Rhenumque anti-
quius transiitibus, propter loci fertilitatem hic con-
fessit. Gallosque, quos loca incolerent, expulsi-
se: solosque esse, qui patrum nostrorum memoria
omni Gallia uocata, Testunet, Cimbrumque intra
fines suos ingressi praesiderent, quae ex re ferri, ut
earum verum memoria magnam sibi aueritatem,
magnaque spiritus in re militari sumerent.

2. IDEM L. V. c. 12. Britanniae pars interior ab
his incolitur, quos mater in insula ipsa memoria pre-
ditum dicitur. Maritima pars ab his, qui, praedat
ac belli inferendi causa ex Belgis transierunt, qui
antei fere in nominibus ciuitatem appellantur, qui-
bus orti ex ciuitatibus ex peregrinis, et bello ille-
gitis his romanis fuerunt, atque agris colere coeperunt.

TACITVS VII. Agricola c. 11. Ratiuae Cal-
doniam inhabitantium Comas, magni artus, Gera-
manicus originem asserunt.

3. CAESAR. L. I. c. 4. Horum omnium
fortissimi sunt Belgae, propterea quod a cultu atque
humanitate prouinciae longissime abstant, minime
que ad eos mercatores seque committunt, atque ea,
quae ad efficiendum animi pertinent, important.
Proximique sunt Germanis, qui trans Rhenum ince-
dunt, quibus cum continentem bellum gerunt.

§. VI. 1. IDEM L. II. c. 4. Condrasus, Ebor-
aci, Caracasi, Pannoni, qui uno nomine Ger-
mani appellantur, & L. VI. c. 31. Segni, Condrusi-
que, ex gente & numero Germanorum, qui sunt in-
ter Eboracum, Treuironem, legatus ad Caesarem
misit, et ait, ne se in bellum numero deceret,
nec omnium Germanorum, qui essent citra Rhe-
num, causam esse unam indiceret.

2. TACITVS de M. G. c. 2. Ceterum Ger-
manis incolentibus rectus & iugis solitudo: quon-
iam, qui primi Rhenum transgressi, Gallos expulerunt
ac nunc Tungri, tunc Germani uocati sunt. Ita
nationis nomen, non gentis euasisse paulatim, ut
vixit primus a uillare ut metum, mox a se ipso,
inuenito nomine, Germani uocarentur. Dicit Cicero,
wie sie an sich bündelt, hat den Gelehrten, die den
Ursprung des Namens Germani erörtern wollen, in
vielen, zum Theil sehr gesungenen Erklärungen be-
halten müssen, denen sie gleichwohl keine alte & ein-
zigste ist. Es erhellet aber doch zum wenigsten so viel
daraus

Von den Namen India vnd America.

India / wirdt eigentlich genandt die große Land-
schaft Afir gegen der Sonnen Aufgang gelegen / so etwann von
Alexandro vnd den Macedonibus besitzet / vnd mit Krieg über-
zogen worden. Wir halten aber dafür / es habe Columbus diesen
Namen India / welcher eigentlich den Morgenländischen zuschre-
bet der Neuen Welt / als der sie erstmals erfinden / darumb gegeben
denn / da er die Neue Welt zuerst erfinden / gemeinet hat / es
sey die Insel Cipangon / welche gegen Sina vnd Catiao über gelegen / mehr der Mor-
genländischen / als Nidergängerischen Grenzen vnd Frontier zugehörig: Vnd dasselbi-
ge nach der Meinung Amiotels vnd Seneca / welche geschrieben haben / India sey
nicht weit von Hispania gelegen.

America aber die Neue Welt / hat von Americo Vesputio / welcher nach Chri-
stophoro Colombo auf des Königs in Castilien Reich Anno 1495. dahin gefogelt
den Namen behalten / vnd sein in drey Theil abgetheilt werden.

Der eine Theil so gegen Nitternacht gelegen / wirdt New Frankreich genandt /
dieweil im Jahr 1524. Johannes Verrazanus ein Florentiner / von König Francisco I.
vnd seiner Mutter / die dem Reich für und zu in die neuer Welt gesandt / das ganze Ge-
stade des Meers gemeret hat / welches sich vom Tropico Cancrj vom 29. Grad / bis
auff den 50. vnd weiter gegen Nitternacht / erstreckt / vnd hat dasselbi des Königs Wa-
pen aufgerichtet / Also / daß die Spanier selbst / die hernach dahin kommen seind / diesen
Theil America / Frankreich genandt haben. Ihre Breite ist vom 15. Grad / bis auff
den 54. gegen Nitternacht / die Länge von dem 280. bis auff den 330. grad. Der Theil
gegen Aufgang / wirdt von den neuen Entdeckten Norumbega genandt / vnd erstreckt
sich bis in den Meerhafen Gamas / damit sie von Canada / (dahin Robertvallis vnd
Jacobus Carterus im Jahr 1535. kommen seind) unterschieden ist. Vnd diese liegen
viel andere Inseln mehr / vnd vnter denselben das Land Labrador / das sich bis gen
Gronelandiam erstreckt. Gegen Nidergang begriffen es viel Landschaften / so nun-
mehr etlicher maffen bekandt seind / als Quivira / Cibola / Astlan vnd Teulichichi-
meni / das aber gegen Nittertag liegt / wirdt Florida genandt / dieweil am Palmtag / den
die Spanier Floridum Paicha nennet / man denselben wahrzunehmen. Der
Theil gegen Nitternacht ist noch gar unbekant.

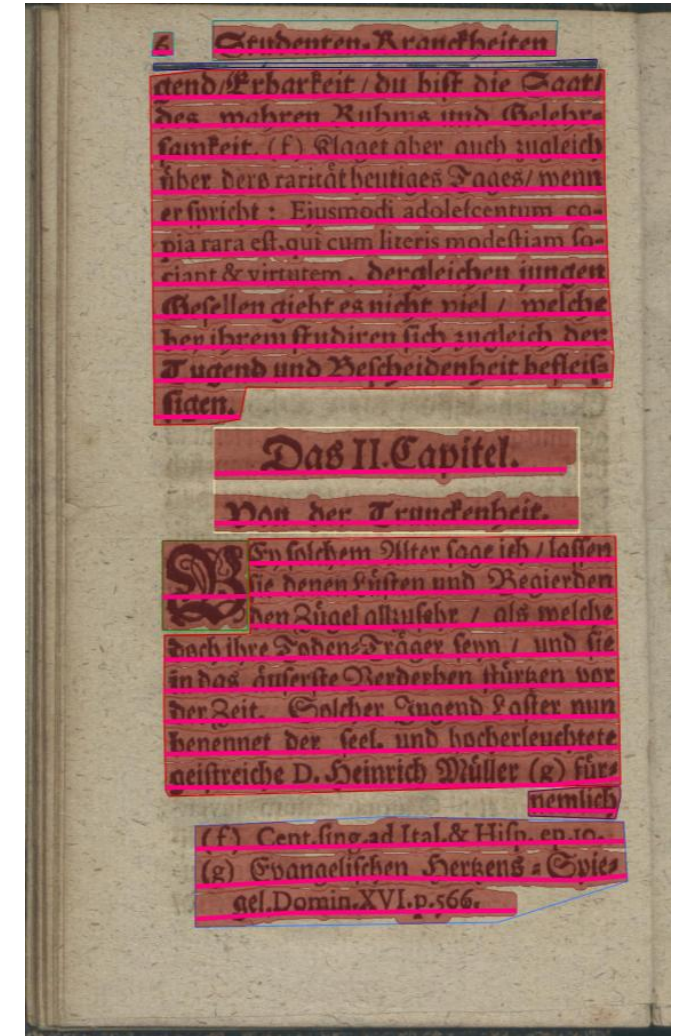
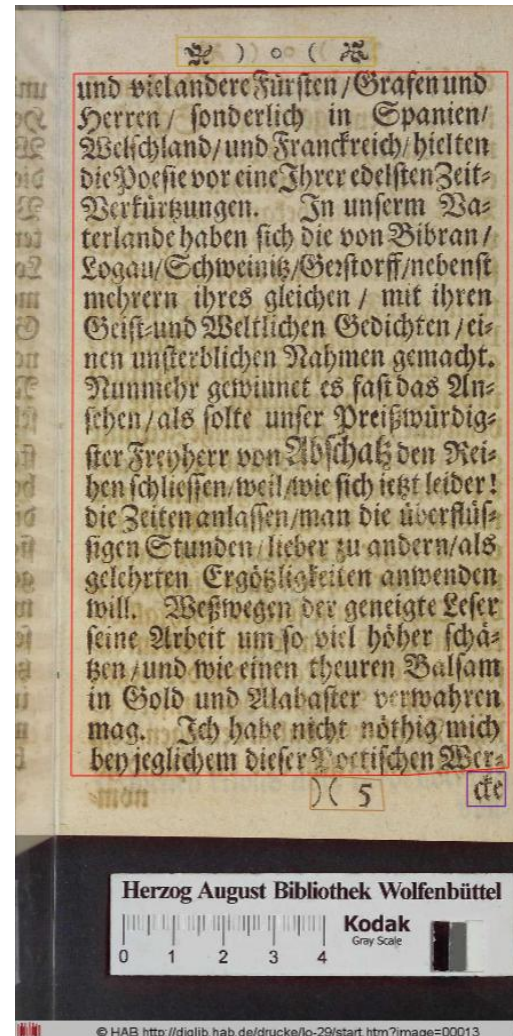
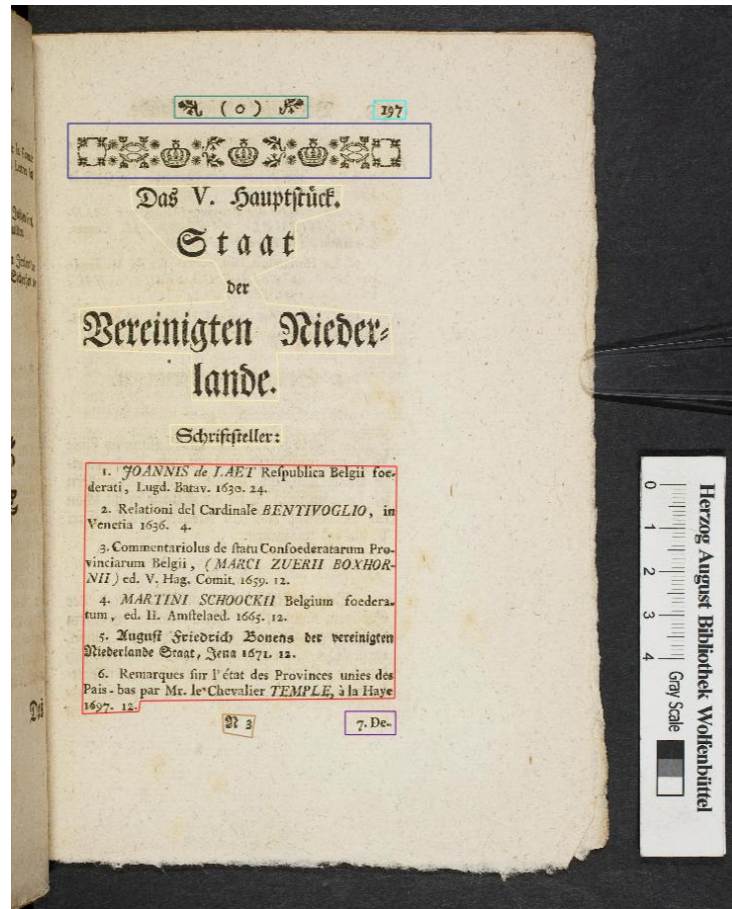
Dieses new Frankreich ist fast so groß als vnter Europa / vnd darinnen die Land-
schaft Florida am besten erbauet / als welchen viel Franckosen in mancheren Schif-
faren angereiset / vnd also die hertzliche ist. Dem Vorgebürg erstreckt sich 100.
Franckische Meilen lang in das Meer / vnd vnter sich findt gegen Nittertag / bargegen
über hat die Insel Cuba 25. Franckische Meilen weit / welche sonst Cubela genand
wirdt. Gegen Aufgang Bahama vnd Lucaia. Gegen Nidergang hat es den Na-
men des Mexicanischen Meers.

Der

Niederlag
America.

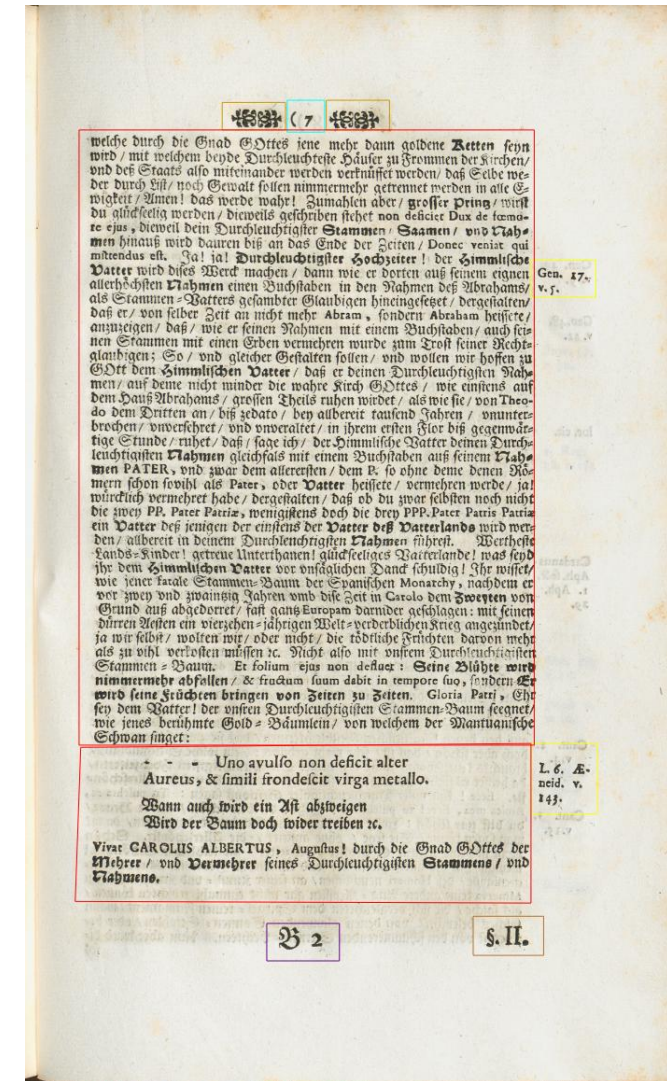
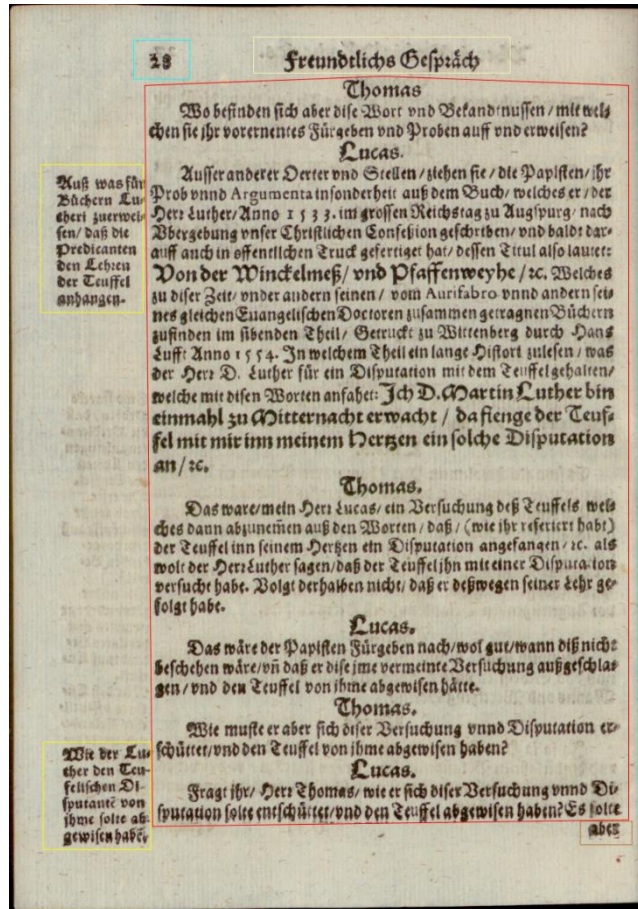
Der erste
Theil Ameri-
ca.

Der zweite
Theil Ameri-
ca.



PRImA

- <https://www.primaresearch.org/datasets>
- PRImA Research Group (Salford, Manchester)
- Pattern Recognition & Image Analysis
- Viele Datensets, meist für Competitions (ICDAR, ...)
- Häufig recht weit vom OCR-D-Fokus entfernt
- Problem: Oft nur kleinere Beispielsätze vorhanden und/oder keine PAGE-XML
- Lösung: Neu ausgezeichnet → insgesamt knapp 400 Seiten



19

Hills, California. And it's thanks to our Mr Garrison's eccentric buying habits that this is now one of the most fascinating Pacers you've ever likely to come across. You see, the car was ordered with so many factory options that he could quite easily have bought a Cadillac for the same money. It's amazing what some people would do back then to save on fuel.

Back in 1976, AMC offered two Pacer models: the basic Pacer X and the D/L, which stood for Distinctive and Luxurious. This particular example is a D/L, coming with numerous equipment gains over the X. Over and above the D/L pack, at the time of ordering the car Lee Garrison added virtually every possible factory option available. The list included air conditioning, tinted windows all round, alloy-style wheel trims, a vinyl roof, quarter light windows, a heated rear window, rear wash-wipe, roof rack and an exterior trim pack – which comprised rubbing strips, over-riders and bumper trims, along with stainless brightwork on the wheel arches and doors.

THE CAR WAS ORDERED WITH SO MANY FACTORY OPTIONS THAT HE COULD QUITE EASILY HAVE BOUGHT A CADILLAC FOR THE SAME MONEY. IT'S AMAZING WHAT SOME PEOPLE WOULD DO BACK THEN TO SAVE ON FUEL.

was no easy task; in fact the body of the Pacer had to be widened by six inches. Later AMC would use this as a plus point in advertising, saying "if you want to ride like a Pacer, you've got to be wide like a Pacer".

In practice, the wideness of the Pacer did give it an amazing ride, and the width – which was actually about the same as a Cadillac – meant interior cabin space that was very good for a compact. AMC also added an interesting "safety feature", which amazes people to this day: the passenger door is four inches longer than the driver's. The idea was that if it were easier to get into the back from the pavement you would be safe from traffic. Fair enough, I suppose.

Initially at least, the Pacer was a reasonable success story for AMC, with 145,528 cars being sold in the first year. Despite it being quite heavily criticised by the press, many American buyers seemed to like the Pacer for its nonconformist looks. However, the novelty wore off quickly and sales went downhill rapidly. The tide was turning, and increasing numbers of Americans were mocking rather than praising the oddball Pacer.

The Pacer's relative lack of power didn't help, thanks to its engine being strangled for economy and emissions. For the 1976 model year, a 120hp 4.2-litre High Output engine with bags of torque was offered as an option, but proved too thirsty for the performance it gave, especially when compared with Japanese imports that were by then prolific. Later for the 1978 model year, a 5.0-litre V8 was offered, but this was too late in the day to save the Pacer – the 'buzz' had long since worn off. After 280,000 units had been produced, the last Pacer finally rolled off the line in 1980.

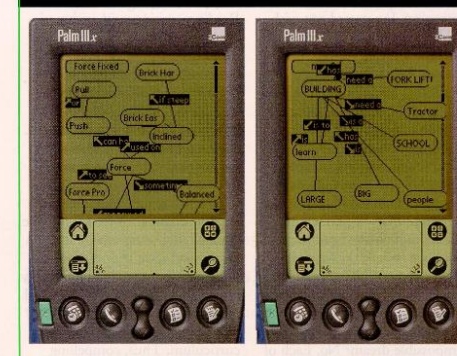
Lee Garrison, the original owner of the amazing Pacer you see here, bought this car new back in 1976, via an AMC dealer in Beverly Hills, California.

Mechanical options ordered by Garrison included the High Output 4.2-litre engine and an uprated suspension package with stiffer springs and a front anti-roll bar. (Having driven this particular car, we'd hate to think what a Pacer without the suspension pack drives like!) This Pacer also came with power disc brakes, limited slip differential, sports automatic transmission with column changes, over-ride, power

29

Log On Education

Figure 1. Examples of PicoMaps created by students.



that enable learning and teaching.

And (here is the really important bit) link those applications to existing curricular materials so educators essentially already know how to use the handheld devices. Having enough applications that have educational utility is strong ammunition in arguing for a switch from graphing calculators, a definite one-trick pony, to handheld devices. Here then, are our candidates for some effective handheld applications.

PicoMap. PicoMap is a concept mapping tool for handheld devices. Figure 1 presents several PicoMaps created by 10- to 12-year-olds in Detroit during a unit on the physics of heavy machinery. PicoMap goes beyond paper-and-pencil concept maps in the

following ways:

- First, children beam their PicoMaps to each other. For example, children will pair up and, in a group brainstorming activity about, say, water quality, one child will identify sources of chemical pollution (nitrates) in their local river while the other child will identify sources of physical pollution (sewage). Then the children beam each other their ideas to support the ensuing discussion.
- Since each child has a handheld device ready-at-hand, children naturally revise their PicoMaps. In contrast, if their documents are on a desktop computer accessed only one hour per week, children will not engage in

the critically important process of iterating and revising their documents.

Children can print out their PicoMaps, allowing the teacher to track what the children are doing and give feedback, while parents use the printouts for refrigerator decorations. Classroom management issues are not to be taken lightly; making printing straightforward makes teachers comfortable with having each of their 30 students equipped with a handheld device. (If only syncing 30 handheld devices to one desktop computer was as simple.)

Palm sheets. For better or worse, worksheets are a fixture in K-12 classrooms. But a handheld device's worksheet has numerous advantages over its paper cousin: the handheld device can immediately check a student's input and provide feedback. After the data from the worksheet are transferred to a desktop computer, the data can be automatically aggregated and graphs presented that depict all the students' answers. Figure 2 shows an 11-year-old filling out an Air Quality Inventory worksheet.

Cooties. How do germs spread? Drawing on the work at the MIT Media Lab with SmartBadges, we developed a socio-kinesthetic simulation on handheld devices to help children understand this process. Children "meet" each other by walking around a classroom with a handheld device and beaming each other either a digi-

FIRST

future. According to Dealogic, 87 SPACs have begun trading in the U.S. since the end of 2003, buying some well-known companies like smoothie purveyor Jamba Juice. Last year alone, 40 SPACs worth \$3.4 billion were announced, up from \$484 million two years earlier.

SPACs are essentially shell companies. They go public with little more to show investors than a management team and an agreement that the money raised will be used to fund an acquisition in a particular sector, such as retail, or in an emerging market like China. Ergo, a SPAC is a roll of the dice. "Imagine paying \$50 to go to a Broadway show, and you have no idea what's behind the curtain," says lawyer Mitchell Littman of Littman Krooks LLP, who works with a lot of SPACs. "You are relying on the fortune and integrity of the management team." Once a deal is completed, the SPAC's managers (who typically receive 20% of the public shares as compensation) are free to sell their holdings, usually after a lockup period.

Today's SPAC boom harkens back to the flurry of "blank check" companies that sprang up in the 1980s. Those were widely discredited by a wave of scams in which fraudsters would take a shell company public, announce a merger, pump the stock, and then dump it before everyone realized that the hot target company was anything but. Investors have a lot more protection with SPACs, which must hold almost all the money in escrow until a deal is done.

From Charney's perspective, there is no downside to the arrangement. It lets him raise cash relatively quickly, remain the largest shareholder, and avoid having to sell other investors on his vision until after his company is already publicly traded. "I'm in a hurry to get the money," he says, citing his plan to expand beyond his current 145 stores. "What's great about the SPAC is, it's kind

of in reverse. First you get publicly traded, then you do the compliance stuff. And if you brought in private equity, they might want to run the company in a more [cautious] manner," he says, referring to American Apparel's risqué image.

Keeping control is particularly important for a guy who is the creative force behind the brand, as well as one whose flamboyant management style—there is one sexual-harassment suit outstanding against him by a former employee; two others have been dismissed—might scare off many investors. "He is very talented," says former J.C. Penney CEO Al



(long after Ledecy cashed out). Boldfaced names on the board include Kerry Kennedy, Bobby's daughter, and Edward Mathias, a managing director at the Carlyle Group. On Jan. 31 the same group formed a new \$250 million SPAC, Victory Acquisition, underwritten by Citi. The filing notes that it could compete directly with Endeavor. And as with Endeavor, once Victory goes public it will pay a Ledecy "affiliate" \$7,500 monthly to run its Manhattan office.

Hedge funds are drawn to these investments. Since SPACs typically must announce a deal within 18 months of their IPOs or return the money raised, with interest, hedgies see them as safe places to park cash while potentially profiting from the price fluctuations between a deal's announcement and its close—a sort of arbitrage play. That may be why Steven Cohen's \$12 billion hedge fund SAC Capital snapped up 8.7% of Endeavor's stock when the deal was announced.

Even if investors are SPAC-happy, not all regulators are. So far only the American Stock Exchange lists SPACs, and many bankers and lawyers say the SEC is deliberately taking its time vetting deals, hoping to slow down the pace of offerings. SPACs are probably best left to those, like Charney, who like to live life on the edge.

UNHEALTHY RETURN

Since smoothie chain Jamba Juice went public in November via a SPAC, its shares have fallen 19%.

FIRST YOU GET PUBLICLY TRADED, THEN YOU DO COMPLIANCE.

SPAC Attack!

The volume of special purpose acquisition companies going public in the U.S. is rising sharply.

Year	Number of deals	Value of deals (in billions)
2004	13	\$0.5
'05	29	\$2.1
'06	40	\$3.4









Fine Wine Shrines

WHILE THE CHATEAUX OF FRANCE HAVE LONG WELCOMED WINE TOURISTS, ENOTURISMO IS only just coming of age in northeastern Spain's Rioja Valley. But what a vintage—the winemakers have hired some of the world's top architects to create full-bodied bodegas to impress. Here are a few that are stopping traffic; be sure to arrange visits in advance.

—BY CLAIRE SMITH

1 YSIOS

Visitors to Ysios' new bodega, set against the spectacular backdrop of the Sierra de Cantabria that shields the Rioja vines from northwesterly winds, may feel like wine pilgrims. Architect Santiago Calatrava has created a cathedral-like building with undulating curves that echo a row of barrels. www.bodegasyisios.com

2 MARQUES DE RISCAL

The residents of Elciego, a small Rioja winemaking town, probably never expected to see Los Angeles-based architect Frank Gehry's trademark titanium curves on their skyline, but there they are. It took 10 years and \$77 million, and Gehry's 43-suite hotel, vinotherapy spa and Michelin-starred restaurant opened at the 146-year-old bodega late last year. "I wanted to design something exciting, festive, because wine is pleasure," says Gehry. www.marquesderiscal.com

3 CVNE

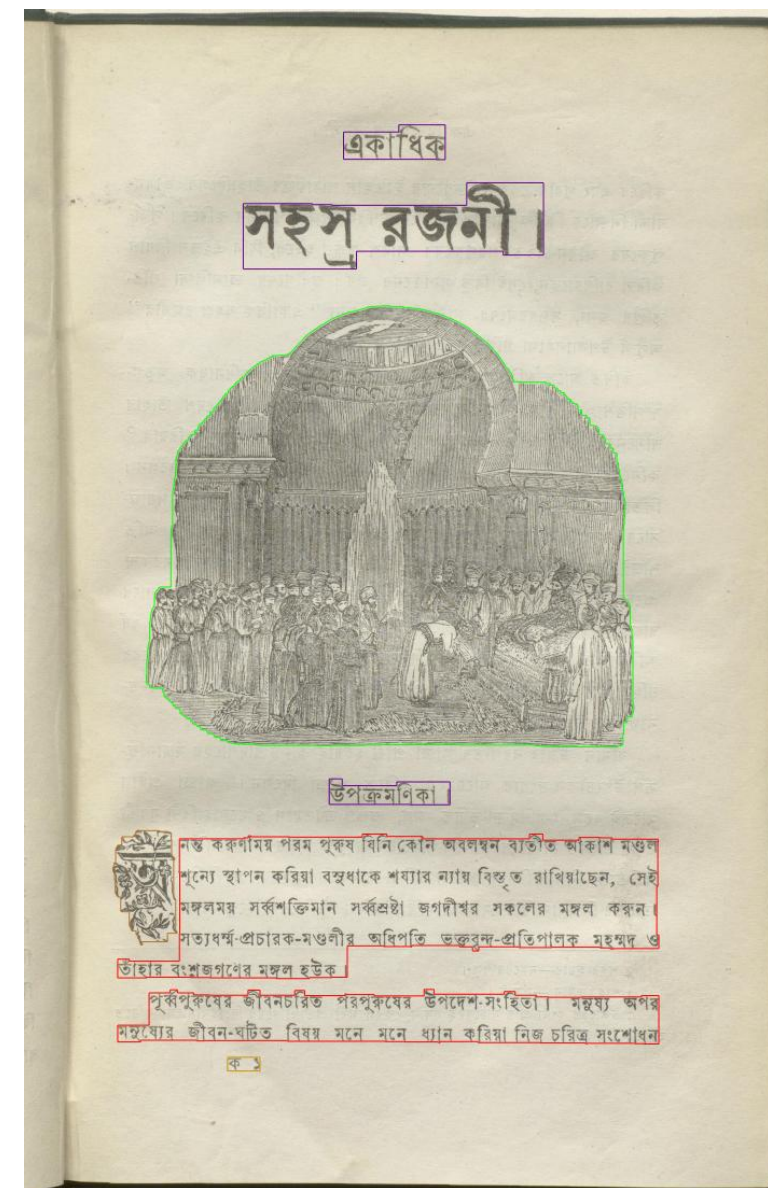
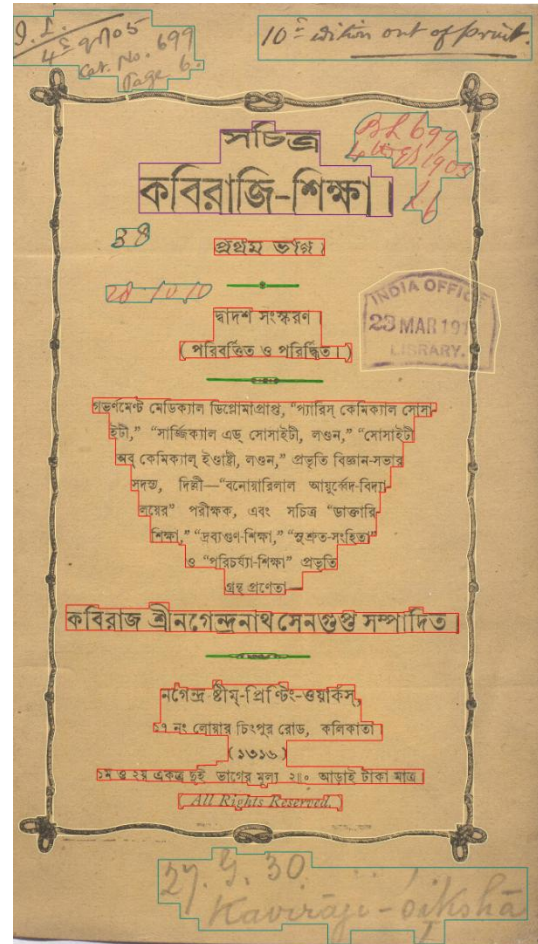
Enter this barrel-shaped bunker and step into French architect Philippe Mazières' vision of 21st century winemaking. Towering 17.5 m above a glass floor, a giant mechanical arm delivers crushed grapes into the 72 stainless-steel tanks that line the circular walls. Even more impressive are two enormous caverns bored into the hillside that store 22,000 barrels of *crianza*. "Making wine used to be like a secret in the Rioja," says tour guide Nunia Noja. Now it's like the underground lab in a James Bond film. www.cvne.com

4 R. LOPEZ DE HEREDIA

"We have always been sensitive to beauty," says María José López de Heredia, a fourth-generation winemaker, who commissioned British architect Zaha Hadid to design a wine-tasting shop for the winery's HQ in Haro. The result: a mirrored white bottle-like building incorporating a carved bar that Heredia's great-grandfather took to the 1910 Universal Exhibition in Brussels. Says López de Heredia, "My great-grandfather designed a bodega 129 years ago, and half of it is to be built still. Our dream is to finish it." www.lopezdeheredia.com



REID2017



REID2019

ভূগোলবিদ্যাসার	
১১। সমুদ্র নদী প্রভৃতি জলাশয়ের তরঙ্গ উৎপন্ন হয়।	তীর
১২। নদী প্রভৃতি প্রোতোযুক্ত জলাশয়ের নিকটস্থ উচ্চ ভূমি তরঙ্গে আহত হইলে তাহাকে কাছাড় বলা যায়।	কাছাড়।
১৩। সমুদ্রাদি জলাশয়ের মধ্যে পর্কত, চর, বা অন্য কোন প্রকার জলমগ্ন স্থাবর অদৃশ্য থাকিলে তাহাকে গ্রস্তিকা বলা যায়। এই স্থানে নৌকাদি বানিচালি হইয়া থাকে, এই জন্য নাবিকেরা তাহাকে জিঞ্জিরা বলে।	গ্রস্তিকা।
১৪। নৌকা বানিচালি হইবার ভয়ে নাবিকেরা যেখানে নিদর্শন চিহ্ন রাখিয়া থাকে, তাহাকে গ্রস্তিকা ধ্বজ বলা যায়।	গ্রস্তিকাধ্বজ।
১৫। জলমগ্ন বালিচড়ার নাম গুপ্তচর। নাবিকেরা তাহাকে মশিনা কহিয়া থাকে।	গুপ্তচর।
১৬। জলাশয়ের উপর দিয়া যে স্থিতিকাদিনাম্নিত গমনাগমনের পথ চলিয়া যায়, তাহার নাম সেতু বা পুল। যথা শ্যামবাজারের পুল, খিদিরপুরের পুল ইত্যাদি।	সেতু।

BENGALI LIBRARY—Catalogue of Books registered in the Presidency of Bengal during the quarter ending the 31st December 1919—concluded.			
Author and title, brief subject, including the age of the book where the name is obscure, number of pages, publisher and place of publication, date, given on the title page with the name of the printer when other than the Calcutta Press—also given from the press or of publication, date, edition and price.		Printer and place of printing.	Number of copies.
BENGALI AND SANSKRIT PERIODICALS—RELIGION—concluded.			
41. হিন্দু পত্রিকা। [Hindu Patrika. The Hindu Magazine, 26th year. No. 6. A monthly magazine devoted to Hindu religion, literature and science. Edited by Jadunath Majumdar, M.A., B.L. Pages 241-288. Published by the printer, Jessore. October, 1919. [12th October, 1919.] 4 th 1st edition.		Kalpeshana Chatterji, Hindu Patrika Press, Jessore.	1,200
Price, 4 annas.			
42. [12th November, 1919.] 4 th 1st edition.		Price, 4 annas.	1,200
43. নিত্যধর্ম বা সর্বধর্ম সম্বন্ধে (নৈমিত্তিক)। [নৈমিত্তিক]। [Nityadharma Va Sarvadharmam Samabandha. A monthly journal devoted to the harmonization of all religions. 6th year. Nos. 9 and 10 (together). A monthly Hindu journal advocating the unification of all religions. Edited by Upendra Nath Mazumdar, M.A., B.L. Pages 129-160. Published by the editor, Malacram Math, Kalighat, Calcutta. Aavin and Kartik. 1326 s.d. or September-November, 1919. [21st November, 1919.] 8 th 1st edition.		Niradharan Das, Titilash Press, 70, Kalighat Street, Calcutta.	500
Price, 4 annas.			
[Nos. 7 and 8 (together) noticed in entry No. 30 at page 150 of the Catalogue for the quarter ending September, 1919.]			
44. সজ্জন ভোবনী (শ্রী)। [Sri Sajjan Toshani. The Comforter of the Pious. 22nd year. Nos. 1 and 2 (together). A monthly journal devoted to Vaishnav religion. Edited by Siddhanta-Saravati. Pages 1-60. Published by the editor, Mayapur, Katia. 438, Banskandol 1918, 20 A.D. [20th September, 1919.] 16 th 1st edition.		Jugendra Chandra Halder, Sri Bhagavat Press, Krishnagar.	500
Price, Re. 1 yearly.			
[21st year, Nos. 11 and 12 (together), noticed in entry No. 33 at page 150 of the Catalogue for the quarter ending September, 1919.]			
45. বৈষ্ণব সঙ্গিনী বা ভক্তি প্রভা (শ্রীমতী)। [Sri Sri]। [Vaishnav Sandini Va Bhaktiprabha. Female Companion of the Vaishnavs or Light of Devotion. Vol. XV. Nos. 2 and 3 (together). A monthly journal devoted to Vaishnav religion. Pages 17-30, 225-240. Published by Mathuradas Das, Bhikari, Bhari, Hinglip, Bhudra and Aavin. 1326 s.d. or August-October, 1919. [22nd September, 1919.] 8 th 1st edition.		Sakhamay Mitra, New Arya Mission Press, 3, Shivnagar, Das Lane, Calcutta.	400
Price, Re. 1-8 yearly.			
[No. 1 noticed in entry No. 36 at page 150 of the Catalogue for the quarter ending September, 1919.]			
46. No. 4. Pages 37-62, 241-248. Kartik. 1326 s.d. or October-November, 1919. 8 th 1st edition.		Price, Re. 1-8 yearly.	400

Schulbücher

- DFG-Projekt „Digitalisierung und Erschließung historischer Schulbücher mit Schwerpunkt religiöse Bildung und Aufbau einer Wissensbasis für die bildungshistorische Forschung“
 - Kooperation zwischen GEI Braunschweig und ZPD Würzburg
 - Förderung in Line „Digitalisierung & Erschließung“
 - Laufzeit 3 Jahre, Start 07/24
- Ziele:
 - Erfassung von deutschspr. Religionsschulbücher und schulischen (Erst-)Lesebüchern
 - Systematische Erfassung von Akteuren aus der historischen Schulbuchproduktion unter Einsatz von Named Entity Recognition (NER) und Named Entity Linking (NEL)
- Zahlreiche Daten erfasst (→ knapp 4.500 S.), viele Trainingsexperimente (gleich mehr)

22. MONARCHIÆ PRIMÆ.

8. Aeneas succediret seinem Schwieger-Vater Latino im Regiment.

9. Er ist der erste König der Latiner von denen/die nach der Zerstörung Troja in Italien kommen.

10. Von dem Todt Aeneas biß auff die Erbauung der Stadt Rom werden 426. Jahr gezehlet.

CENTURIÆ TERTIÆ. Decas I.

1.

Thineus
XXX.

Thineus regieret 30. Jahr / von An. M. 2831.

2. Die Labe des Bundes wird von den Philistern weggenommen. 1. Sam. 4. v. 11.

3. Samuel succediret dem Eli An. M. 2850. und ist so wohl ein Prophet als ein Politischer Richter gewesen.

4. Dem Ascanio Aeneas Sohn folget im Regiment nach Sylvius, von welchem hernach die Succesores Sylvi genennet werden / weil er in einem Walde gebohren.

5. Die Heraclidæ kommen in den Pelopponesum, und nehmen Spartam, Argos und Mesfenen ein/welches geschehen 80. Jahr nach der Verwüstung Troja.

Dercylus
XXXI.

6. Dercylus regieret 40. Jahr von A. M. 2861.

7. In



109. Samische Amphora mit Negertanz. 6. Jahrh. (Nach Hoerber. Griechische Vasen.)



110. Attische Amphora mit dem Zweikampf des Diomedes und Aeneas. 5. Jahrhundert. (Nach Robert.)

Die Malerei tritt in unserer Vorstellung vom klassischen Altertum nur deshalb hinter den beiden Schwesterkünsten, Architektur und Plastik, zurück, weil sie in vergänglicheren Stoffen schuf; in Wirklichkeit hatte sie zeitweise sogar die Führung. Immerhin können wir die Wandlungen der großen griechischen Malerei an der Hand der hochentwickelten griechischen Vasenfabrication verfolgen. Der wichtigste Fortschritt ist der von der schwarzfigurigen zur rotfigurigen Technik. 109 zeigt einen Doppelhenkelkrug, dessen Bauch ein drolliger Negertanz ziert; das sinnvolle Flechtband am Hals des Gefäßes erinnert an die Säulenbasen des Erechtheion (16). Die Figuren sind mit glänzend schwarzem Firnis auf den roten Ton aufgetragen. Die rotfigurige Malerei, welche gegen Ende des 6. Jahrhunderts v. Chr. die schwarzfigurige ablöst, spart die Figuren auf dem roten Tongrunde aus und gewinnt so die Möglichkeit einer feineren Innenzeichnung. Die farbige Tafel II zeigt den Zweikampf zwischen Diomedes und Aeneas im Anschluß an Ilias V, 239 ff., nur daß der Felsstein, mit dem Diomedes seinen Gegner an der 1. Hüfte verwundet, durch die künstlerisch besser darstellbare Lanze ersetzt wird.

Eine Probe griechisch-römischer Wandmalerei gibt die farbige Tafel III. Sie zeigt den sog. zweiten Dekorationsstil, welcher die Wand durch gemalte Pilaster und Säulen zwar immer noch architektonisch gliedert, aber sie doch auch schon scheinbar durchbricht und so ideell raumerweiternde Ausblicke ins Freie schafft. Die Wand ist (wie in 107 das Mittelfeld der Rückwand durch wirkliche Säulen) dreigeteilt. Das Mittelfeld ist nach Art eines Tempelchens eingerahmt und eröffnet unter einem Stichbogen den Blick auf eine mythologische Szene aus der Jugend des Dionysos. In die beiden Seitenfelder sind scheinbar Tafelbilder auf Marmor eingelassen. Auch die übrige Dekoration zeigt bei freiestem Spiel der Phantasie erlesensten Geschmack und feinsten Farbensinn; dem festlichen Zinnoberrot halten die kühleren bläulichen, grünlichen und schwarzen Töne das Gleichgewicht.

Die Kunst der Mosaikmalerei erreichte im Altertum eine hohe Blüte. Ihr Ursprung ist darin zu suchen, daß man bunte Steine in den gestampften Estrich einlegte, zunächst in geometrischen Mustern, bis man schließlich, zuerst im hellenistischen Osten, namentlich in Ägypten, auch wirkliche Gemälde in Stein übertrug. Ein solches Gemälde ist der Kentauren- und Raubtierkampf aus der Villa Hadrians in Tivoli (Tafel IV). Der überaus lebendig heranspringende Kentaure wirft einen Felsblock auf den Tiger, der seine Krallen in den Leib der tot am Boden liegenden Kentauren eingeschlagen hat. Einen Löwen hat er bereits erlegt, ein Panther bedroht ihn von einer Felsplatte aus. Die sehr harmonische Farbenwirkung ist auch hier auf den Gegensatz und die Vermittlung von warmen und kalten Tönen gestellt.

Groberungspläne zu fördern, und ließ Truppen in Polen einrücken. Um nicht das ganze Land eine Beute Rußlands werden zu lassen, schloß Friedrich Wilhelm II., wie sein großer Vorgänger, mit der Kaiserin einen Teilungsvertrag, wonach große Stücke des östlichen Polens an Rußland, Danzig, Thorn und die von nun an als Südpreußen bezeichneten Gebiete an Preußen fielen. Österreich war an dieser Teilung nicht beteiligt; um so eifriger beobachtete es die Vergrößerung Preußens.

Jetzt entstand in Polen eine starke nationale Bewegung zum Schutze des zerstückelten und in seinem Dasein bedrohten Vaterlandes; an ihre Spitze trat Kosciuszko. Aber trotz verzweifelter Widerstandes erlagen sie dem russischen General Suworow, der Praga, die Vorstadt Warschans, erstürmte und die Hauptstadt darauf einnahm. Nunmehr folgte die dritte Teilung Polens, das aufhörte ein selbständiger Staat zu sein. Preußen erhielt Neuostpreußen mit Warschau, Österreich 1795. Westgalizien, Rußland die großen Gebiete, die übrig blieben.

§ 10. Der erste Koalitionskrieg. Bonaparte. Die wichtigsten Schlachtplätze des ersten Koalitionskrieges waren Belgien, die Rheinlande und Oberitalien. Der Organisator der französischen Armeen war Carnot, welcher Mitglied des Wohlfahrtsausschusses war. Sie wurden auf Grund einer allgemeinen Aushebung gebildet, die unter den jungen Leuten vom 18. bis zum 25. Jahre stattfand; so brachte die französische Revolution wieder den Grundsatz der allgemeinen Wehrpflicht zur Geltung, während die Heere der Verbündeten aus Berufsoldaten bestanden. Ohne sich an die Regeln der überlieferten Kriegskunst zu lehren, gingen die kühnen, jungen Generale der Franzosen frisch auf den Feind los. Verluste konnten sie durch neue Aushebungen leicht ersetzen; ihren Unterhalt beschafften sie sich durch Requisitionen. Bald waren sie fast überall im Vorteil. Das linke Rheinufer mußte von den Verbündeten geräumt werden; ebenso eroberten die Franzosen Holland und verwandelten diesen Staat in eine batavische Republik.

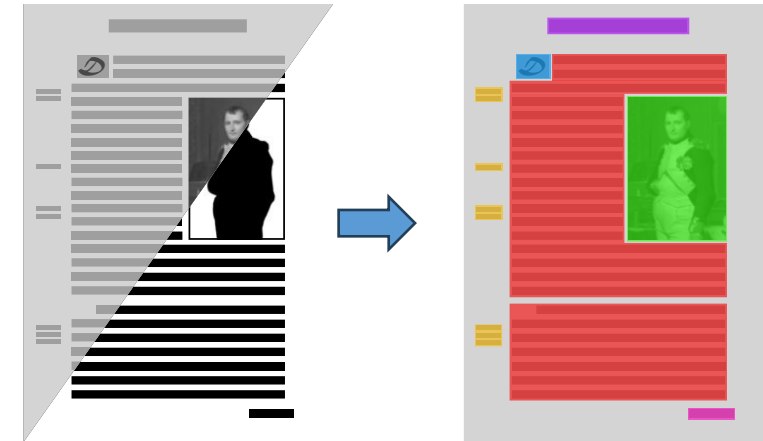
Da entschloß sich Friedrich Wilhelm II., teils weil ihm die Geldmittel zur Fortführung des Krieges fehlten, teils weil seine Truppen durch die polnischen Ereignisse in Anspruch genommen waren, zum Frieden mit Frankreich, der zu Basel im Jahre 1795 abgeschlossen wurde. Durch eine Demarkationslinie wurde Norddeutschland von dem süddeutschen Kriegsschauplatz geschieden; dort herrschte Friede, hier tobte der Krieg weiter. Während die Waffen über das Schicksal Europas entschieden, be-

Agenda

1. Projektübersicht
2. LAREX (Teaser)
3. Trainingsdaten
- 4. Trainings-Testereien (BA Janik Haitz)**
5. (Meine) Fragen

Übersicht

- Hier: Fokus auf die Identifikation und Segmentierung von Regionen, Zeilensegmentierung dann quasi trivial
- Kann je nach Material und Benutzeranforderungen sehr anspruchsvoll sein
 - Text \leftrightarrow Nicht-Text
 - ... (viele Zwischenstufen)
 - Fließtext \leftrightarrow Überschriften \leftrightarrow Marginalien \leftrightarrow ...
- Ziel für das Schulbuchprojekt: eher feinkörnige, aber pragmatische und sinnvolle semantische Unterscheidung
- Verwendete Software: Kraken (<https://kraken.re>)
 - Pixel Classifier + Postprocessing Heuristiken
 - Eingabe: Bild einer einzelnen Seite; Ausgabe: Bereichspositionen und -typen (PAGE XML)



Trainingsbeispiele

paragraph

page number

— 135 —

— 187 —

135

Bald sieht man von ferne die Türme der Paläste und Tempel der Hauptstadt, in der der Pharao Hof hält. Freilich,* zuerst gilt es die äußeren Bezirke der Stadt zu durchheilen und da ist nicht viel Schönes zu sehen: kleine, armselige Lehmhütten, dazwischen unregelmäßige, schmale Gäßchen oder schmutzige Plätze, oft auch ein Tümpel, in dem Ochsen baden. Das niedere Volk ist sehr arm, es fristet nur eben sein Dasein und kann seine Hütten offen stehen lassen, da es dort nichts zu stehlen gibt.

Vorstadt.

Kommt man dagegen näher zum Mittelpunkt der Stadt, so wird es allmählich anders. Die Häuser sind größer und manchmal recht hoch, so daß man in den engen Gassen kaum den Himmel sieht. Dann öffnet sich ein Platz, auf dem ein Markt abgehalten wird. Die Leute drängen sich eifrig handelnd und feilschend durcheinander und haben für nichts anderes Augen und Ohren. Und der Handel ist auch manehmal gar nicht einfach. Eigentliches Geld gibt es nicht, sondern nur Edelmetallstücke, die bei jedem Handel neu abgewogen werden müssen, wobei man immer Angst hat, übervorteilt zu werden; daher ziehen viele den Tausch vor. Hier bietet der eine selbstverfertigte, dauerhafte Sandalen zum Tausch für einen Topf mit wohlriechender Salbe, dort werden Perlenketten gegen Weihrauch und andere Wohlgerüche verhandelt. An einer anderen Stelle wird um einen Stier gefeilscht und man muß für ihn 1 Matte, 5 Maß Honig, 8 Maß Öl und noch 7 kleinere Dinge geben. Das muß man alles genau in Metallwert ausrechnen und dann erst kann der Tausch durchgeführt werden. Vom Markte gehen enge Gassen nach verschiedenen Seiten aus. Hier wohnen Kaufleute und Handwerker: Schuster, Bäcker und Zuckerbäcker, Garbküche, Goldschmiede, Tischler; aber auch Wirtstuben gibt es, in denen Wein und ein Gerstengebräu, unserem Bier ähnlich, geschänkt wird. Auch Waren aus den asiatischen Kulturländern werden hier verkauft und ein gar buntes Treiben herrscht, fast wie auf dem Markte.

Die Stadt.

Handel.

marginalia

Gewerbetreibende.

Hofburg.
Auszug des
Pharao.

Nun weiten sich die Gassen und nach einiger Zeit stehen wir vor der königlichen Hofburg, die sich fast wie eine Stadt in der Stadt ausnimmt. — Im Augenblicke, da wir uns nähern, tritt die Spitze eines Zuges aus dem Palast hervor. Voraus eilen Läufer, die die Aufgabe haben, das Herannahen des Pharao zu melden. Alles Volk weicht schon zur Seite, es öffnet sich ein weiter freier Raum. Nun folgt der Zug: an der Spitze eine Abteilung auserwählter Leibgarden, dann Fahnenträger, andere Leute, die große Wedel, Zeichen der königlichen Würde, tragen, dann Krieger, die mit Lanzen, Beil und Schild gerüstet sind, und endlich der Pharao selbst. Allein steht er auf seinem Wagen, in ein langes, wallendes Gewand gehüllt; wie ein Gott erscheint er dem Volke, das sich vor ihm auf den Boden wirft. Hinter ihm kommt die Königin, gleichfalls im wallenden Gewand, und dann die Prinzen und hohen Beamten. Der König wird von Sklaven geleitet, die große Wedel über sein Haupt halten. So zieht er in den Tempel, um dort seinem Gott, als dessen Sohn er sich betrachtet, ein Opfer darzubringen.

Ist will i dein Nachfolger sein.
Herr! machs nach deim Verlangen.

3. Ein Pilgram bin i halt monneht,
Muß rafa fremde Strofa,
Das bitt i di, mein Gott und Herr
Du wirst mi nit verstoja.

4. Der Glaubn hob i frei bekennt,
Das dorf i mi nit schäma,
Wenn mo mi glei ein Kexer neunt
Und tut mir's Leba nehma.

5. Muß i glei in das Elend fort,
Bil i mi do nit wehra,
So hoff i do, Gott wird mir dort
Och gute Fründ Leichera.

6. Herr, wie du wilt, so gib mi drein,
Bei dir wil i verbleiba,
I will mi gern dem Willn dein
G'duldig unerschreiba.

7. Mueß i glei fort, in Gottes Nam,
Und wird mir alles genomma,
So wuß i wol, die Himmelstren
Wer i emohl bekomma.

8. So mueß i hent von meinem Haus,
Die Kinder! mueß i laffa,
Mein Gott, es treiß mir Jährel aus
Zu wandern fremde Stroja.

9. Mein Gott, fähr mi in eine Stadt,
Wo i dein Wort kann heba,
Darin will i die früh und spä
In meinem Herzen loba.

10. Sol i in diesem Jammeral
Noch länger in Armut leba,
So hoff i da, Gott wird mir dort
Ein beßre Wohnung geba.

52. Türken und Franzosen

Lut, jetzt ist es dahin kommen,
Daß du werdest ganz laseth,
Dann wir haben's vorgenommen,
Daß dein Reich zu Ende geht.

Swar der arge Schandfranzose,
Der die Welt auskäuferint.

image region

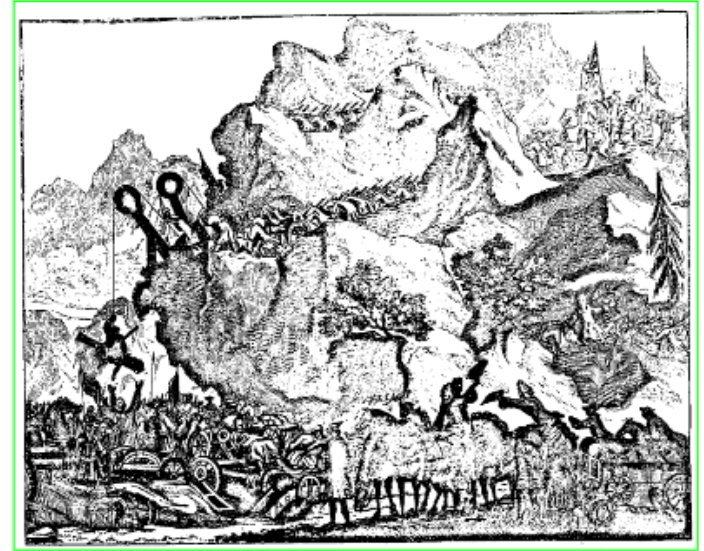


Fig. 43. Der „sehr mühsame und wunderwürdige Marsch“ der kaiserlichen Armee über die Tirolischen und Alt-Veronesischen Alpen im Jahre 1791.

caption

Anschläge der Gegner zu vereiteln. Ludwig XIV., mit seinen bisherigen Feldherren unzufrieden, übergab den Oberbefehl an Villeroi (sprich: Willroa), der versprochen hatte, er werde Eugen tot oder lebendig dem Könige überliefern. Aber Eugen überrumpelte den Prahler zur Nachtzeit in der Mitte seiner Truppen und führte ihn, ehe die Seinigen ihm Hilfe bringen konnten, als Gefangenen fort. Selbst in Paris sang man Spottlieder auf Villeroi. Auch gegen Villerois Nachfolger Vendôme (sprich: Wandohn).

¹⁾ Südöstl. von Verona; nicht zu verwechseln mit der Stadt Carpi südlich vom Po.

footnote

separator
region

heading

Experimente


■ Annotierte Datasets

- Training: 800 gemischte Seiten; spezialisierte Sub-Datasets: 202 Titelseiten, 348 ToC-Seiten
- Evaluation: 241 gemischte Seiten; 25 Titelseiten und 25 ToC-Seiten

■ Testen verschiedener Setups (weitestgehend analog zur Texterkennung)

- Effekt der Input-Bildformate: grau \geq Farbe \gg binär
- Pretraining und Datenaugmentierung hilft, allerdings nur ein bisschen

■ Ergebnisse

- Sehr präzise, speziell auf normalen Inhaltsseiten: Pixel Accuracy (PA) von 95,4% 
- Seitentypen-spezifische Modelle (Titel, ToC, ...) führten zu deutlichen Verbesserungen
 - PA (Titelseiten): 89,0% \rightarrow 96,1%
 - PA (ToC-Seiten): 87,6% \rightarrow 97,7%

Haitz, Janik (2025): *Implementation and Evaluation of Methods for Layout Analysis in the Context of Automatic Text Recognition of Historical Schoolbooks*. Bachelorarbeit.

Beispielsergebnisse I

78 79 Der Frühjahr-, Sommer- und Herbstfeldzug 289

Während des Waffenstillstandes schlossen Rußland und Preußen ein Bündnis mit England. England versprach Subsidien zu zahlen, Preußen dagegen nach dem Friedensschluß Ostfriesland an Hannover abzutreten. Auch Schweden, wo der bei Napoleon in Ungnade gefallene Bernadotte zum Nachfolger des kinderlosen Königs gewählt worden war (1810), trat dem Bunde bei. Alexander hatte ihm für Finnland Norwegen zugesagt, das er nach dem Kriege den Dänen zu entreißen beabsichtigte.

Wahnis mit England.

Österreich unternahm es jetzt, zwischen den Kämpfenden zu vermitteln. Metternich verhandelte mit den Abgesandten beider Parteien in Prag; er ging dabei lediglich von dem Interesse seines Staates aus und wünschte ihn von Napoleon zu befreien, aber zugleich zu vermeiden, daß an Stelle des französischen ein russisches Übergewicht trete. Seine Vorschläge ließen Napoleon ein mächtiges Frankreich übrig, wurden aber von ihm zurückgewiesen. Kurz vor Ablauf des Waffenstillstandes erklärte daher Österreich seinen Beitritt zur Koalition.

Österreich.

§ 79. Der Sommer- und Herbstfeldzug. Napoleon hatte sein Heer inzwischen auf etwa 450 000 Mann verstärkt und links der Elbe zusammengezogen; sein Hauptquartier befand sich in Dresden. Eines seiner Heere stand in Spanien gegen Wellington im Kampfe (vgl. § 70); überdies erforderten die ausgedehnten Küsten seines Reiches bedeutende Kräfte zum Schutz gegen einen Landungsversuch der die See beherrschenden Engländer. Von den Rheinbundstaaten hielt Bayern einen Teil seines Kontingentes noch zurück. Die vereinigten Heere der Verbündeten waren etwa 480 000 Mann stark. Sie hatten drei Armeen aufgestellt: 1. die Nordarmee (etwa 126 000 Mann) unter Bernadotte, dem Kronprinzen von Schweden, mit dem Hauptquartier in Berlin, bestand aus dem schwedischen und zwei preussischen Korps (unter Bülow und Tauentzien); 2. die Schleifische Armee unter Blücher mit Scharnhorst als Generalstabschef stand mit zwei russischen und einem preussischen Korps (unter York) etwa 104 000 Mann stark bei Breslau; 3. die Böhmisches oder Hauptarmee (254 000 Mann), die sich um Prag sammelte, bestand aus dem gesamten österreichischen Heere, den russischen Gardes und einem preussischen Korps (unter Kleist) stand unter dem Befehle des Fürsten Schwarzenberg mit Radetzky als Generalstabschef. Die drei verbündeten Monarchen besaßen sich in Schwarzenbergs Hauptquartier. Nach dem gemeinsam festgestellten Plane sollten die Armeen gegen die Aufstellung des Feindes gleichzeitig vorgehen, dagegen jede einzelne zurückweichen, wenn der Feind überlegene Kräfte gegen sie wende.

Die Aufstellung.

Die Armeen.

Kriegsplan.

Nach Ablauf des Waffenstillstandes ging Napoleon zum Angriff gegen Blücher vor und entsandte Dubinot gegen die Nordarmee; diesen sollten Truppen aus Magdeburg und Davout aus Hamburg unterstützen. Als Dubinot am 23. August aus dem wald- und fumpfreichen Gelände zwei Meilen südlich von Berlin herausrat, wurde er bei Großbeeren von den preussischen Korps unter Bülow und Tauentzien angegriffen und besiegt.

Nordarmee.

Großbeeren (23. August).

Preiser, Geschichte. VI. G. (2b.)

9

280 Namen- und Sachverzeichnis.

- Breisach 54. 106, 1.
Breitenfeld 50.
Bremen 23. 54. 130.
Breslau 23. 147.
Bretonischer Klub 178, 2.
Brieg 134. 136.
Brief 65.
Brienne 191.
Brüggemann 17.
Brüssel 232.
Brumaire, der achtzehnte, 196.
Buchdruckerkunst 14.
Bülow 216. 218.
„Bürgerkönig“ 232.
Bürgerliche Verfassung der Geistlichkeit 179.
Bulgarien 276.
Bulletin, das 29., 213.
Bundesakte 221.
Bundesrat 235, 1. 257. 268.
Bundestag 221. 239. 255.
Bunzelwitz 152.
Burgheyl 85.
Burgund 73.
Burke 167.
Burkersdorf 153.
Burschenschaft 225. 226.
Butler 52.
Byron 229, 2.

C.
(Vergleiche K. und Z.)

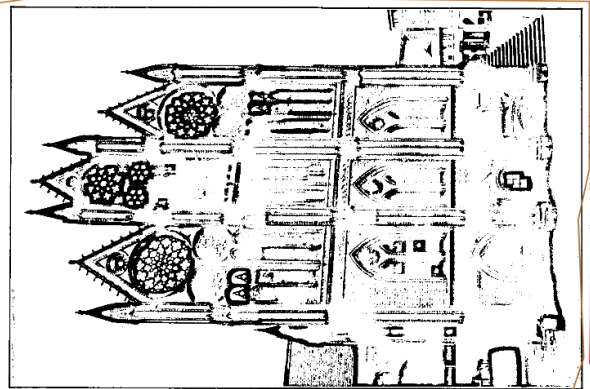
- Cabral 9.
Cadiz 208.
Cadoudal 198, 3.
Cajetan 18.
Calais 62. 72. 82.
Calderon 70.
Calonne 174. 175.
Cambrai 28.
Camoës 8, 1. 70.
Campo Formio 192.
„Canaille“ 57.
Canning 229.
Capri 274, 6.
Carbonari 227.
Carlos, Don, 61.
Carlos, Don, 143.
Carlos, Don, 234.
Carnot 189.
Cartesius 4.
Câteau-Cambrésis 62.
Cavaignac 236.
Cavour 247. 248.
Cecil 85.
Censur 226.
Centraljunta 207. 208.

- Centrakommission in Mainz 226.
Centralverwaltungsdepartement 217, 4.
Cervantes 69.
Cevennen 115.
Châlons-sur-Marne 262.
Chambord 232, 1.
Charleroi 104, 2.
Châtillon 218.
China 8.
Chotusitz 149.
Christian II. von Dänemark 119.
Christian III. von Dänemark 32.
Christian IV. von Dänemark 46. 47.
Christian VIII. von Dänemark 239.
Christian IX. von Dänemark 240. 253, 4.
Christine von Schweden 51. 122.
Christine von Spanien 234.
Chrysoloras 13.
Churchill 110.
Cisalpinische Republik 192.
Cisleithanien 258.
Clemens VII. 28.
Clemens XIV. 164.
Clive 170, 2.
Code Napoleon 199.
Colbert 113. 114.
Coligny 74. 75. 76.
Colombey 262.
Common prayer book 83, 1.
Condé, Ludwig von, 74. 75.
Condé, Ludwig von, 53.
Confessio Augustana 29.
Cook 13.
Corday 184.
Corneille 116.
Cornwallis 169.
Cortes, Ferdinand, 10. 11.
Cortes, die, 61. 208.
Cotta 18.
Courbière 203.
Couthon 185.
„Covenant“ 90. 93.
Cranmer 81. 82.
Crespy 30.
Cromwell, Oliver, 92—94.
Cromwell, Richard, 94.
Cuba 9. 10.
Custozza 244. 248.
Cuzco 11.

D.

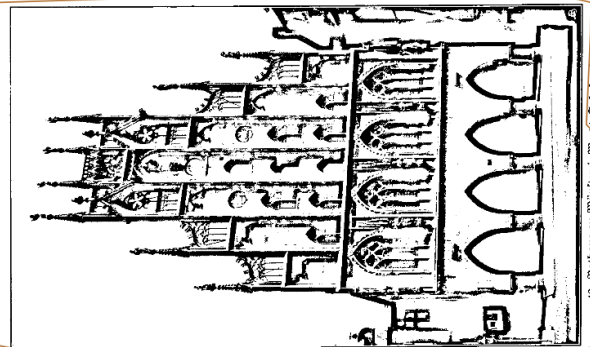
- Dänemark 32. 118, 9. 123. 218. 253, 4.
Dalmatien 193.
„Damenfriede“ 28.
Danton 183—185.
Danzig 204, 3.
Darnley 84.

22 IX. Die gotische Baukunst: Rathausbau.



43. Rathaus zu Langenmünster (Vor. v. H. Schuch Berlin).

42 zeigt die Anwendung der Gotik auf den Profanen. Das gotische Haus kehrt auf langem schmalen Grundriss die hohe Giebel-front der Straße zu. Unten der dem öffentlichen Verkehr dienende belebte Laubengang, dahinter die Wohnräume, oben die mit reichem Schmuck ausgestatteten Gemächer des Patriciats. Der treppenförmig abgetufelte Giebel weist eine andere Teilung auf, in welche sich die Giebelöffnungen pyramidenförmig einordnen. Das Mäuerwerk des Giebels klingt in Kantengehänge von Staffeln zu Staffeln hinunter. In der Giebelöffnung stellen sich die Kirchen und Burgen der Rathhäuser ebenbürtig zur Seite. Stolz stehen sie die den eigentlichen Giebel bei weichen den profanen Gebäuden. Der Giebel über den Dachstuhl empor, um weit ins freie Land hinaus zu wirken. Das Mäuerwerk der Rathhäuser ist aus Formsteinen zusammen; die materielle Glanzung des ganzen Baues wird durch bunte Glanzung wirkungsvoll. Giebelung bei 43 von unten bis oben durch.



44. Rathaus zu Wittenberg (Vor. v. C. Schuch Berlin).

Beispielsergebnisse II

mixed model

ToC model

Inhalt ¹⁾.

I. Mittelhochdeutsche Zeit.

1. Epik.

A. ~~Drinks~~-Grape.

Das Ribelungenlied. a) Inhalt. (Von Gudrun u. Holfand)	3
b) Bruchst. Der Abschied der Burgunden von Pechlarn. (Übersetzt von Karl Barfisch)	14
Gudrun. a) Inhalt. (Nach August Wilmar)	17
b) Bruchst. Ein Vogel bringt Kunde. (Übersetzt von Wilhelm von Plönies)	23

B. Kunst-Epoe oder höfische Epoe.

a) Inhalt. (Von Dr. Alexander
 Schöppner) 26
 b) Buchbild, Parzivals Entstehung. (Übersetzt von San Marte) 28

2. Höfische Lyrik oder Minnegefang.

Mittelfänger

Walthar von der Vogelweide. 1. Morgengebet. (Übersetzt von Karl Simrod)

2. Reinmars des Alten Tod. (Übersetzt von Adalbert Schröter)

3. An Leopold von Österreich. (Übersetzt von Adalbert Schröter)

4. Das Fest zu Wien. (Übersetzt von Karl Simrod)

5. Drei Höfe. (Übersetzt von Karl Simrod)

6. Erziehung. (Übersetzt von Karl Simrod)

Reinmar von Zweter. Güter Wunsch zu gutem Ende. (Übersetzt von Friedrich Rückert)

3. Didaktik.

Aus „Freibanks Befreiendheit“. (Übersetzt von Adolf Barmeister) 36

II. Neuhochdeutsche Zeit.

1. Die Reformationszeit (16. Jahrhundert).

Kaiser Maximilian I. Leuerdant. Inhalt. (Nach Karl Goebcke)	38
Martin Luther. Ein Kinderlied auf die Weigenachten vom Kindlein Jeſu	41
Von der Muſica	43

1) Die mit * bezeichneten Lesestücke sind in der früheren Auflage nicht enthalten.

Inhalt ¹⁾.

I. Mittelhochdeutsche Zeit.

1. **Эпик.**

A. Volks-Epos.

	Seite
D as Nibelungenlied. a) Inhalt. (Von Ludwig Uhland)	3
b) Bruchstück. Der Abschied der Burgunden vom Reich. (Übersetzt von Karl Vofsiß)	14
G udrun. a) Inhalt. (Nach August Wilmar)	17
b) Bruchstück. Ein Vogel bringt Kunde. (Übersetzt von Wilhelm von Blönnies)	23

B. Kunst-Epos oder höfisches Epos.

Parzival, von Wolfram von Eschenbach. a) Inhalt. (Von Dr. Alexander Schöppner)	26
b) Buchstüd. Parzivals Erziehung. (Übersetzt von San Marte)	28

2. Höfische Lyrik oder Minnegefang.

Minnefänger.

Walter von der Vogelweide. 1. Morgengebet. (Übersetzt von Karl Simrod).	
2. Reinmars des Alten Tod. (Übersetzt von Albalbert Schroefer).	
3. An Leopold von Österreich. (Übersetzt von Albalbert Schroefer).	
4. Das Fest zu Wien. (Übersetzt von Karl Simrod).	33
5. Drei Hölle. (Übersetzt von Karl Simrod).	34
6. Erziehung. (Übersetzt von Karl Simrod).	34
Reinmar von Zweter. Güter Wunsch zu gutem Ende. (Übersetzt von Friedrich Rückert).	35

3. Didaktik.

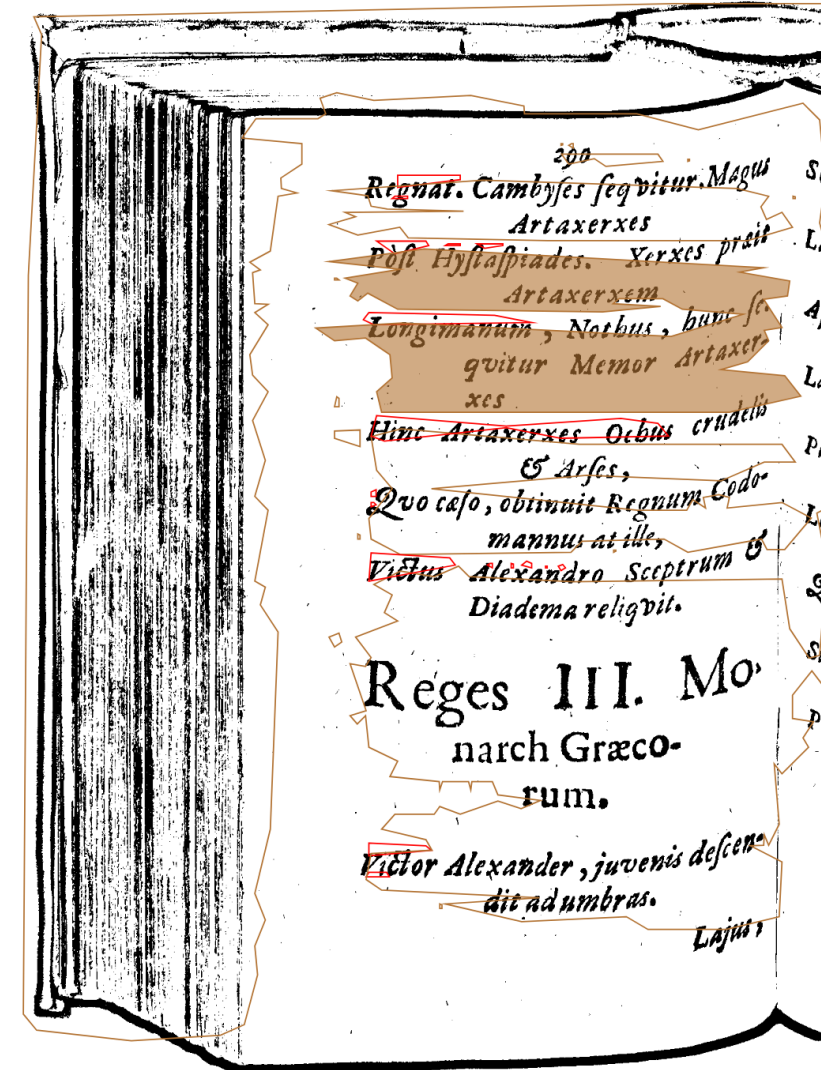
Aus „Freibanks Bescheidenheit“. (Übersetzt von Adolf Barmeister) 36

II. Neuhochdeutsche Zeit.

1. Die Reformationszeit (16. Jahrhundert).

Kaiser Maximilian I. Teuerdank. Inhalt. (Nach Karl Goedeke) . . .	30
Martin Luther. Ein Kinderlied auf die Weihnachten vom Kindlein Jesu . .	41
Ron der Musica	44

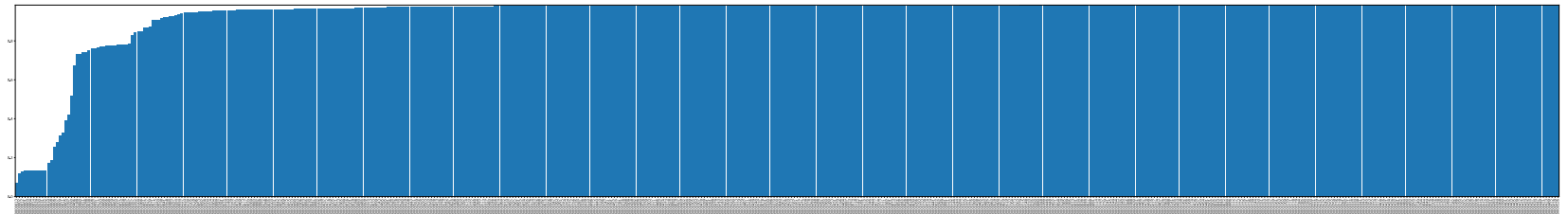
1) Die mit * bezeichneten Lesestücke sind in der früheren Auflage nicht enthalten.
Lesebuch f. L.-H.-N. III. 27



Zusammenfassung

- Layouterkennung auf Großteil der Seiten sehr präzise
- Weitere Verbesserung durch mehr Trainingsseiten (QA → AL!)
- Seitentypen-spezifische Modelle vielversprechend
- Segmentierungsrichtlinien: weniger ist mehr?!
- Multimodaler Ansatz naheliegend und vielversprechend

1. Die Geschichte der Stadt...	107
2. Die Geschichte der Stadt...	108
3. Die Geschichte der Stadt...	109
4. Die Geschichte der Stadt...	110
5. Die Geschichte der Stadt...	111
6. Die Geschichte der Stadt...	112
7. Die Geschichte der Stadt...	113
8. Die Geschichte der Stadt...	114
9. Die Geschichte der Stadt...	115
10. Die Geschichte der Stadt...	116
11. Die Geschichte der Stadt...	117
12. Die Geschichte der Stadt...	118
13. Die Geschichte der Stadt...	119
14. Die Geschichte der Stadt...	120
15. Die Geschichte der Stadt...	121
16. Die Geschichte der Stadt...	122
17. Die Geschichte der Stadt...	123
18. Die Geschichte der Stadt...	124
19. Die Geschichte der Stadt...	125
20. Die Geschichte der Stadt...	126
21. Die Geschichte der Stadt...	127
22. Die Geschichte der Stadt...	128
23. Die Geschichte der Stadt...	129
24. Die Geschichte der Stadt...	130
25. Die Geschichte der Stadt...	131
26. Die Geschichte der Stadt...	132
27. Die Geschichte der Stadt...	133
28. Die Geschichte der Stadt...	134
29. Die Geschichte der Stadt...	135
30. Die Geschichte der Stadt...	136
31. Die Geschichte der Stadt...	137
32. Die Geschichte der Stadt...	138
33. Die Geschichte der Stadt...	139
34. Die Geschichte der Stadt...	140
35. Die Geschichte der Stadt...	141
36. Die Geschichte der Stadt...	142
37. Die Geschichte der Stadt...	143
38. Die Geschichte der Stadt...	144
39. Die Geschichte der Stadt...	145
40. Die Geschichte der Stadt...	146
41. Die Geschichte der Stadt...	147
42. Die Geschichte der Stadt...	148
43. Die Geschichte der Stadt...	149
44. Die Geschichte der Stadt...	150
45. Die Geschichte der Stadt...	151
46. Die Geschichte der Stadt...	152
47. Die Geschichte der Stadt...	153
48. Die Geschichte der Stadt...	154
49. Die Geschichte der Stadt...	155
50. Die Geschichte der Stadt...	156
51. Die Geschichte der Stadt...	157
52. Die Geschichte der Stadt...	158
53. Die Geschichte der Stadt...	159
54. Die Geschichte der Stadt...	160
55. Die Geschichte der Stadt...	161
56. Die Geschichte der Stadt...	162
57. Die Geschichte der Stadt...	163
58. Die Geschichte der Stadt...	164
59. Die Geschichte der Stadt...	165
60. Die Geschichte der Stadt...	166
61. Die Geschichte der Stadt...	167
62. Die Geschichte der Stadt...	168
63. Die Geschichte der Stadt...	169
64. Die Geschichte der Stadt...	170
65. Die Geschichte der Stadt...	171
66. Die Geschichte der Stadt...	172
67. Die Geschichte der Stadt...	173
68. Die Geschichte der Stadt...	174
69. Die Geschichte der Stadt...	175
70. Die Geschichte der Stadt...	176
71. Die Geschichte der Stadt...	177
72. Die Geschichte der Stadt...	178
73. Die Geschichte der Stadt...	179
74. Die Geschichte der Stadt...	180
75. Die Geschichte der Stadt...	181
76. Die Geschichte der Stadt...	182
77. Die Geschichte der Stadt...	183
78. Die Geschichte der Stadt...	184
79. Die Geschichte der Stadt...	185
80. Die Geschichte der Stadt...	186
81. Die Geschichte der Stadt...	187
82. Die Geschichte der Stadt...	188
83. Die Geschichte der Stadt...	189
84. Die Geschichte der Stadt...	190
85. Die Geschichte der Stadt...	191
86. Die Geschichte der Stadt...	192
87. Die Geschichte der Stadt...	193
88. Die Geschichte der Stadt...	194
89. Die Geschichte der Stadt...	195
90. Die Geschichte der Stadt...	196
91. Die Geschichte der Stadt...	197
92. Die Geschichte der Stadt...	198
93. Die Geschichte der Stadt...	199
94. Die Geschichte der Stadt...	200
95. Die Geschichte der Stadt...	201
96. Die Geschichte der Stadt...	202
97. Die Geschichte der Stadt...	203
98. Die Geschichte der Stadt...	204
99. Die Geschichte der Stadt...	205
100. Die Geschichte der Stadt...	206



95

merken sein, das Un'rige mit eigen' Mogen zu überleben, und
und nicht auf Andere verlassen. Des Herrn Auge hilft mehr, als
seine beiden Hände. Eine kleine Nachlässigkeit bringt oft ein gro-
ßes Unglück. Weil einst ein Fußnagel fehlte, so ward das Fuß-
eisen verloren; durch Verlust des Fußes ging das Pferd ver-
loren; durch Verlust des Pferdes ging der Weiler verloren; er
ward ergriffen und erschlagen, weil er nicht für den Fußnagel
geforget hatte.

Zum Fleiß und zur Aufmerksamkeit müssen wir noch die
Sparamkeit hinzufügen, wenn unser Fleiß glücklichen Erfolg haben
soll. Man kann, wenn man die Kunst zu sparen nicht versteht,
die Nase Tag und Nacht über seiner Arbeit haben, und doch als
ein Bettler sterben. Durch ganz Indien ist Spanien nicht reich
geworden, weil es mehr ausgab als einnahm. Kort alle mit den
gelfressenden Narrheiten, Puzschacht, Federri u. s. w. Eine Spar-
keit kostet mehr zu unterhalten, als zwei Kinder, und Sammet
und Seide auf dem Leibe lösen das Feuer in der Küche aus.
Kaufe nur immer, was du nicht brauchst, so wirst du bald ver-
kaufen müssen, was du brauchst. Viele sind arm geworden da-
durch, daß sie Bries um ein Spitzgelo eingelauft haben.

Bei dem Allen aber glaubet nicht, daß es mit eurem Fleiße,
eurer Sparamkeit und Klugheit allein genug sei. Alle diese Dinge
sind vortheilhaft, aber nicht ohne den Segen Gottes. Darum
mühet ihr Gott den Herrn fürchten und fleißig zu Ihm beten und
barmherzig und mitleidig gegen die Armen sein. Denn die Gottes-
furcht ist aller Weisheit Anfang, und wer da bittet, dem wird ge-
geben, und wer die Bräuer liebt, der gefüllt dem Herrn wohl.

122. Ende gut, Alles gut.

Ende böse, Alles böse. Du fangst wohl an, ist gut; du en-
digst wohl, ist besser. Du fangst übel an, ist böse; du endigest
übel, ist noch böser. Wer's gut trifft in beiden Stücken, der hat
den besten Ruhm. Doch wenn ja eins sein soll, will ich lieber
mit Paulo übel anfangen und gut endigen, als mit Judas gut an-
fangen und übel endigen. Gott ist zufrieden, wenn man am Ende
vergütet, was im Anfang verdorben. Das Ende bringt die Krone.
Bei guten bis in den Tod, so will ich dir die Krone des Lebens
geben. Offenb. 2, 10. Gott gebe sie mir und dir, das wir beide
davon tragen unseres Glaubens Ende!



Agenda

1. Projektübersicht
2. LAREX (Teaser)
3. Trainingsdaten
4. Trainings-Testereien (BA Janik, Schulbücher)
- 5. (Meine) Fragen**

(Meine) Fragen

- Wo und wie Daten veröffentlichen?
 - Zenodo?
 - Teils fehlende Bildrechte → Nur PAGE + Verweis?
- Eigene Erfahrungen, speziell
 - zum Einsatz von Trainingspipelines
 - zu one-fits-all / zur Übertragbarkeit von breiten Modellen (Segmentierungsrichtlinien!)
- Weitere open-source Lösungen?
- Weitere lohnenswerte Datensets?
- ...????