

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

Christian Reul und Robert Sachunsky

Forschungsstelle DACHS, Universität Würzburg
SLUB Dresden

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



A screenshot of a digital manuscript viewer. The page contains French text and a large illustration of a person on a horse. The text is annotated with numbered circles (1-14) pointing to specific words or lines. The interface includes a vertical sidebar with document thumbnails and various editing tools at the top.

Annotations on the page:

- Ip font de quelles ie me tais. O pombes folz belletres qui de robes ceulo qui nob
et ne pas su paix a manger et alabenture q
osent demander de honeste ses vies
gues, porcnes desies, labos, auers
desa compre devant cestuy qui nous crea
7 Des conditioz courroux 2 inauiaisties des feines
Gt collago
8 que
9 Pliage
10
11
12
13
14
- Quels ont fait a quelz infamie
A la paix des 4 sans forme
Car luy ont toutz lez se bons nre
Monte deus po es de fance
- Des conditioz courroux 2 inauiaisties des feines
C. Distincte aines chevaux brevet
C'est que mons eut femme
Au moyen de q
Monte deus po es de fance
- Des conditioz courroux 2 inauiaisties des feines
Quels ont fait a quelz infamie
A la paix des 4 sans forme
Car luy ont toutz lez se bons nre
Monte deus po es de fance
- Des conditioz courroux 2 inauiaisties des feines
Ricardes folz e
flourdi z boutz
connoissoient les
mamaines de
feines. Autz feines appo
chez de: So eut done ma
tre. Mes vertes folz, c
criptz, vont brevet des feines
touz lez dce grise longez
autal que des anges blus/
et le leu naturez de soul
es, to estessoz die sonnet se
bages primit matinasse et
tousz conditioz. Mais c except
ne pouzue differer / pourzat
faust q
e chapeine des fein
feines. Vie salye, car de leur
mamaines ie ne me nre
le Joy eut mez agmeray tout
que le autrez chevaux de tou
les bons feines mais pze
les mameas en tous mes
exemptz etemps. Lement
la hure de leur gache has
- Autz feines eut q
e fete tute sa grise blu e fageuse avec ses bonnes mameas rebode par tout
les bonnes e soteray, mais n'eblouirap les mameas tute au monde a v
moy. La femme qui eut alaice de la mamee de noblesse, eut ne se except
ge cez plus grise tricler que bauoit toutes les richesses du monde, elle ne se except
point facilement de vices ne pechez. Le este feine est courroux, son mary
facilement elle multiquera los courroux. Nous sime la trop offusice q
e auant nra
de mettre a occulades hebeup. Louz bestier leust teste chose laquelle estoit reg
te du pays elle vint au roy a lui pze quil luy doulcet otteoy: la puce. Il luy ot
roya, et elle luy pna q
e resuoir la sentence. Fauste chose fit, q
e le moron elle gat
ba le peuple de telle mort. Sabale efort moult rebete dasal de dasal, telle trop

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
 - Flexible 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/structur_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)

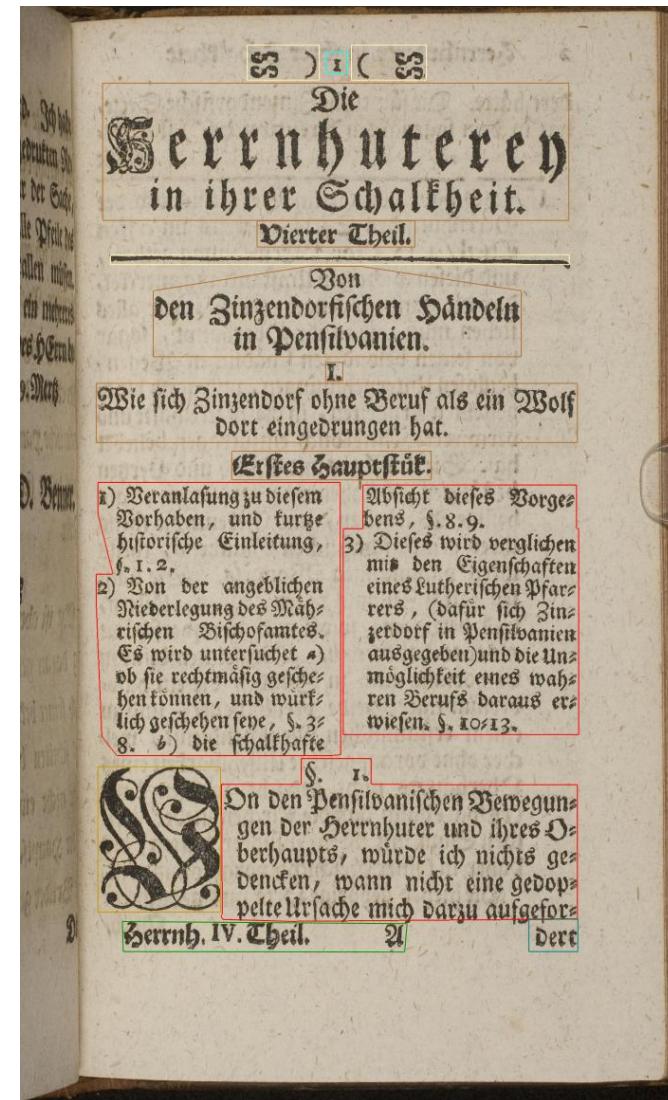


OCR-D Referenzdaten

- <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/bertsky/ocrd_cis
 - (Semi)Automatische Qualitätsanalyse angedacht



DTA – Aussortiert

6

Erfstes Buch. Geschichte der Deutschen

Sohn in den Elfischen Feldern vorher verlündiget.⁶ Auf einer Münze des Corneischen Hauses finden wir diesen Marcellum abgebildet, wie er die Waffen, so er dem erschlagenen Bödmar abgenommen, in dem Tempel des Iouis Fererii aufstellt.⁷ Nur ist es zweifelhaft, ob diesjenigen, so die Fastos Capitolinius abschrieben, recht gelesen, indem unter allen Historicis keiner die Deutschen nennt, und der Name Germani zu der Zeit wohl noch nicht bekannt gewesen. Vielleicht haben sie Germanis, anstatt Germania gelesen: diese waren ebenfalls ein Gallisches Volk, so nedig in Insubrien das Land, wo nachmals Brescia, Cremona und Piacenza angelegt worden, bewohnt, und so wohl an diesem, als nach mährischen Kriegen gegen die Römer Thiel gehabt. Wie wir denn in eben diesen Fällen finden, daß nicht lange nach Marcello, Cethegus über beyde zugleich triumphierte.⁸

V. Deutlicher Nachrichten finden wir von den Colonien, so die Deutschen an jene Seite des Rheins nach Gallien geschickt. IVLIVS CAESAR⁹ berichtet, daß die Belgen, so den dritten Theil von Gallien innen hatten, und zu seiner Zeit die tapfersten unter den Galliern waren, alle Deutscher Abkunft gewesen. Die Treviorer und Noriker insbesondere wußten sich viel damit, daß sie von Deutschen abstammten. Die Belgen haben hernach die Küsten von Britannien besucht¹⁰ und beständig mit den Germanis zu schaffen gehabt.¹¹

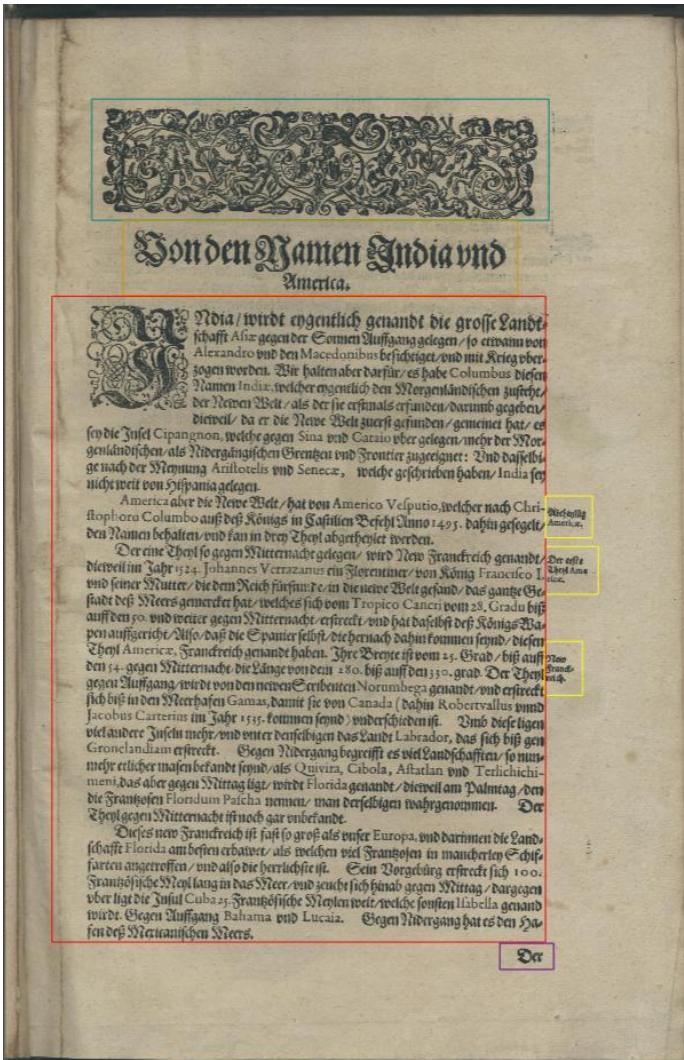
VL. Selbst der Name Germani ist an der andern Seite des Rheins zu erst

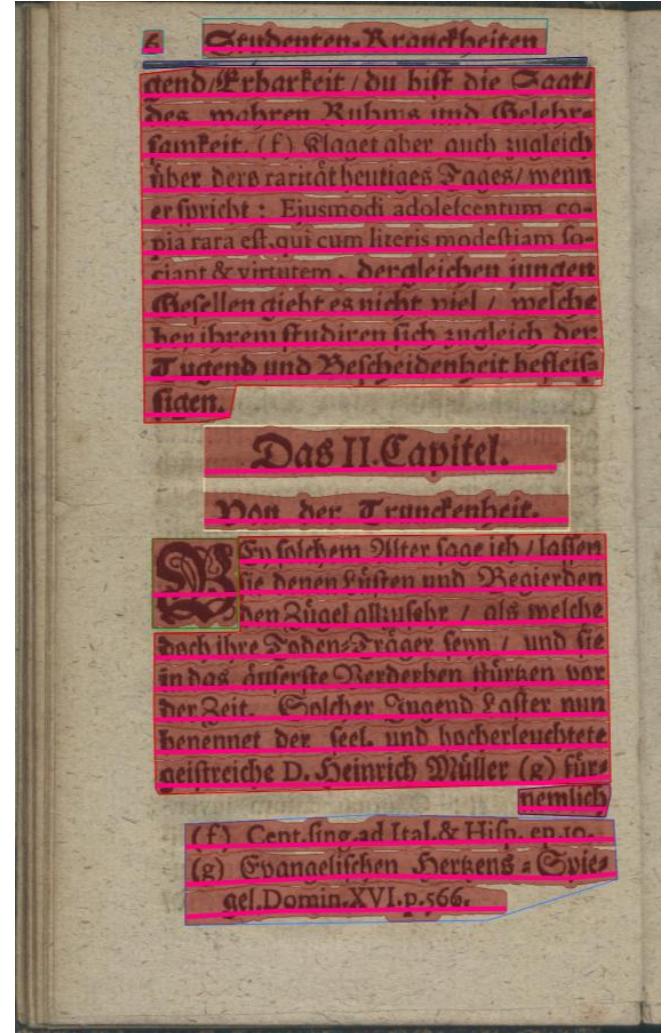
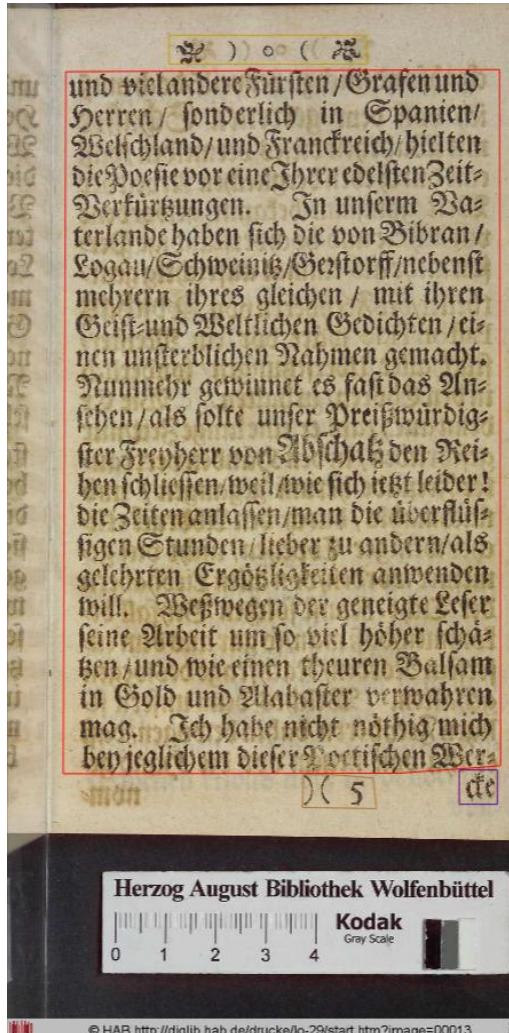
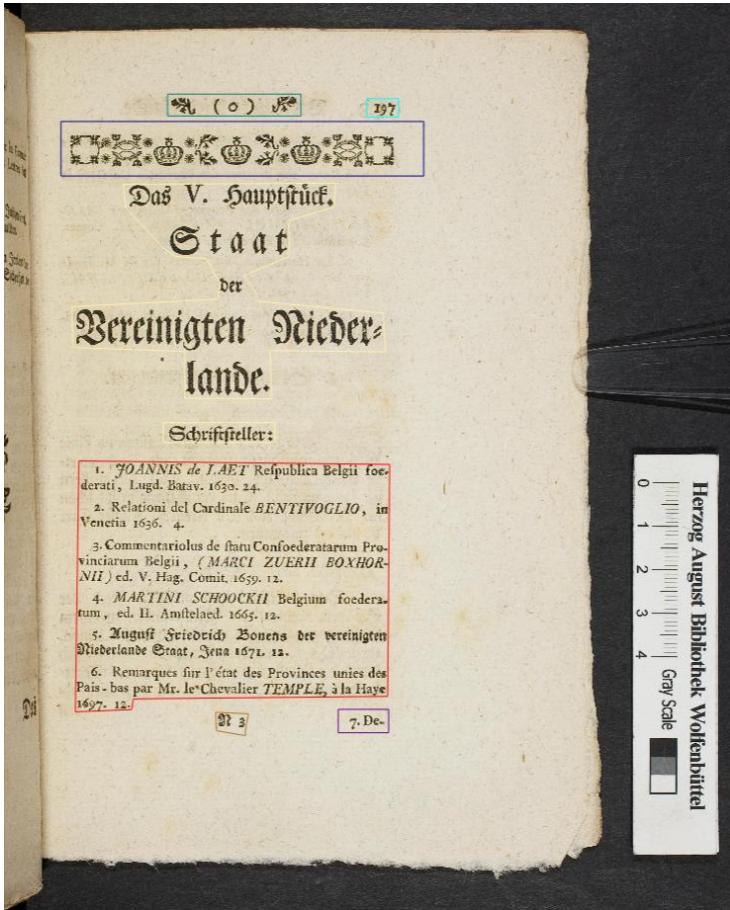
Colonien der
Deutschen
sich in Gallien
aufgegliessen.

6. VIRGILIVS AEN. VI. 255-260.
Apros ut insignis gloria Marcelli opinis
Ingerendus, uiderque viru supereminent omnes.
Hic rem Romanam magna turbatrum ruit.
Sister eques, fratresque Pannonum rebiles
Taurorum pars, superbus opta Quirines.
Confer PERITONIVM in annalibus Hilt. c. 7
7. V. PATINI familiae Romanae in num
p. 87. num. 3-4.
8. Ad an. V. C. DLVI. C. CORNELIVS
L.P.M.N.CETHEGV.COS. DE INVIRBVS ET CENO
L. CALPVS. A. CL. L. TIBERIVS. Q. PETRVS.
C. MARCVS. G. C. L. Q. S. C. M. L. C. L.
C. MARCVS. Q. MARCVS. C. MARCVS.
C. MARCVS. C. MARCVS. C. MARCVS.
que Belga eis orti a Germani, Rhenum anti
quius transduces, propero loci fertilitatem ibi con
fecti; Gallesque, qui ea loca incollerent, exal
fati: falsaque, qui patrum nostrorum memori
omni Gallia uexata, Taurorum, Cambriisque intra
finis horum ingredi problemat, et quod est, ut
earum remanent nomen, magis filiorum vestitum,
mentis pueris, non militari, solum.

2. TACITVS of M.G.C. 2. Ceterum Ger
manicorum sunt recens, & nuper nominati: quin
am primi Rhenum transire, Gallos expulerat
et nunc: Tongri, nunc Germani ecclesiastici, ita
nationi nomen, non gentis exanimi pallidum, ut
omnes primi a ultra ob metum, nunc a scissis
inueniuntur, Germani vocantur. Dicit Etelle,
ac bello inferendo causa ex Belgis transire, qui
enim ferre in nominibus interuersi apparet, qui
bus, ut ex cunctis, sicut etiam & huiusmodi, ex
ibis remanserunt, atque agros colere cooperarunt.
TACITVS Vit. Agricola c. 11. Rutilius Cal-

darus

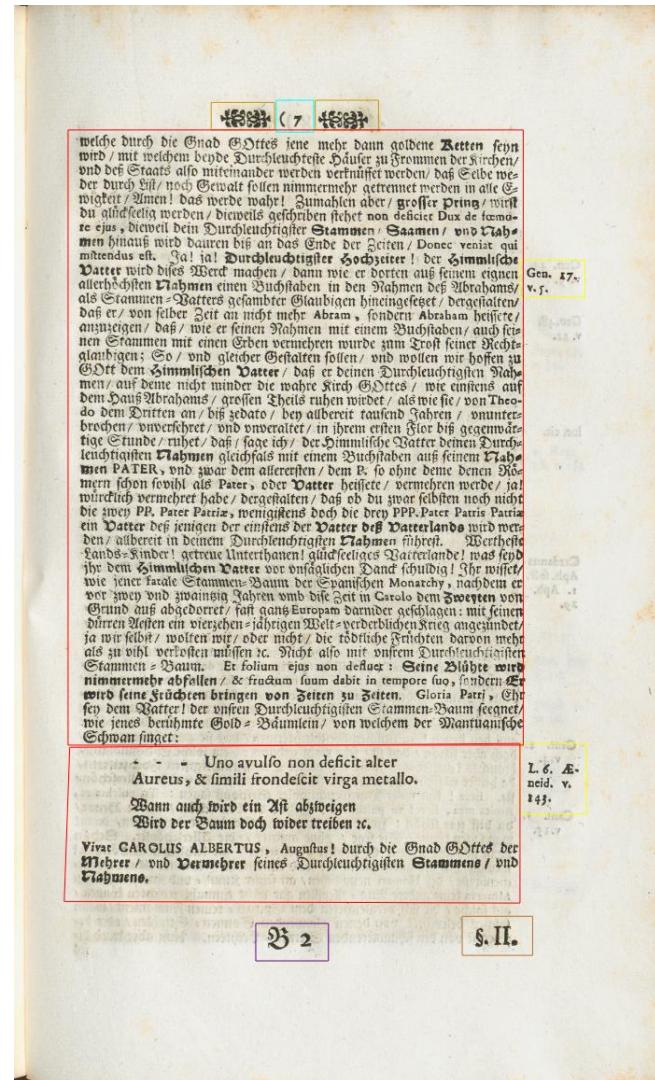
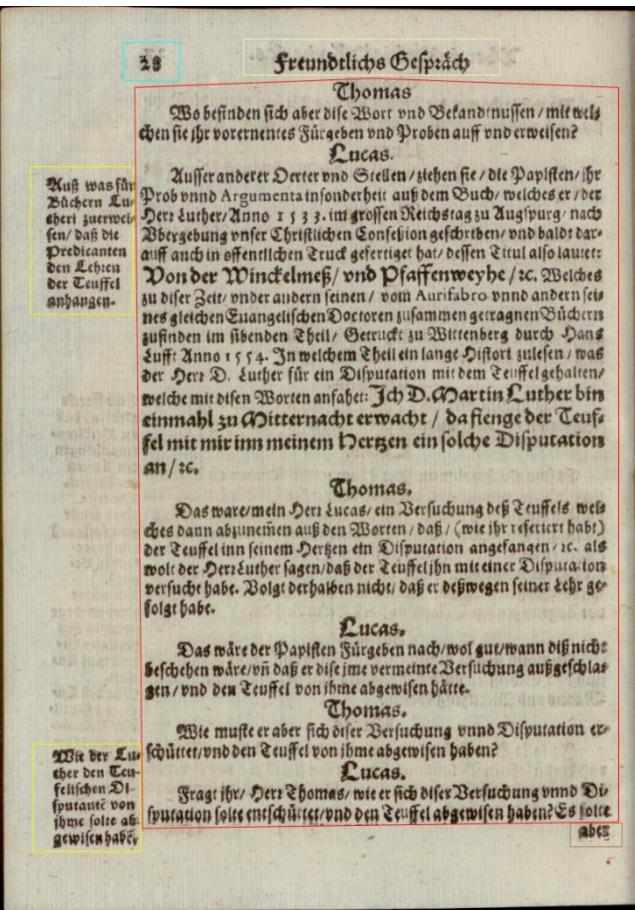




PRImA

- <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 Beispielsets vorhanden und/oder keine PAGE-XML
- Lösung: Neu ausgezeichnet → insgesamt knapp 400 Seiten





Neulengbacher Zeitung

Beilagepreis:
für Neulengbach K 2
Sankt Peter K 2
Duisburg K 2
Krefeld K 2
Mitteleuropa K 6
Gesamtpreis K 6
Ausgabe K 6
Auflage 10000

Unabhängiges Wochenblatt für Politik, Land- und
Sachwirtschaft, Gewerbe, Kunst und Literatur.

Erscheinet jeden Samstag. —

Schriftleitung und Verwaltung: Neulengbach Nr. 35.
Postamt: Neulengbach. —

Auges. Nummer 14. Seiten. 700

Nr. 5. Samstag, den 1. Februar 1902. 5. Jahrgang.

[In dem eingetretenen Kaiserwahljahr richten wir an die geheime Kabinett- und Räte unseres Blattes die Bitte, die Beweisung des Wunsches umgehend vorzuholen zu wollen, damit in der Amtstafel des Kaiserlichen Reichsverbandes erscheint.]

Zur Reichsraths-Eröffnung.

Am 4. Februar wird das überzeugende Rechtswissen eines Abgeordneten mehr als eine lange Sitzzeit, gekört an „Recht und Sitz“, an eine frischgewogene Arbeit nicht genügt, um die von der Regierung ergriffenen und noch ungelösten Verhältnisse aufzuklären.

Die Beweisung ist am 1. Februar im Reichsrathssaal, der Kanzlei des Kaiserlichen Reichsgerichts, in den Sälen des Reichstags, im Deutschen Reichstag, wo nicht wieder der große Zusammenschluss im Regie-Saal.

Aber der Widerstand ist über, es hat an einer politischen Arbeit in den letzten Jahren kein geringeres Interesse als der obige Bericht des Reichstagschefs, der sich auf die harten Fragen des Kriegs kommt. Es fehlt nur die gesetzliche Anerkennung durch den Kaiser, denn der Kanzler hat der Reichsregierung keinen Vorschlag gemacht, um die Beleidigungen, die die Reichsregierung dem Kaiserlichen Reichstag, dem Reichstag und dem Reichstagsgesetz beigebracht hat, mit einer Gesetzesänderung zu bestätigen, ausgenommen die verhältnismäßige Anerkennung der vertraglichen und durch den Kaiser auf Beobachtung der Gesetze erzielten Erfolge, die genau lobenswert ist. So darf alle die politische Arbeit, die im Deutschen Reichstag und im Reichstag verhandelt und nach mir aufgegriffen, die Jahre, welche zu einer zufriedenstellenden Ausbildung der Deutschen und Deutschen führt.

Leuilletton.

As dem Tagebuch einer schlechten Erzieherin.

Sehr eigne Erziehungen von St. Maximilians.

Es war ein trüber Judentag, als ich mich einen Wurm mit „A“ dem jungen Herrn geschickte. Ich habe diesen Wurm mit „A“ gespielt, als ich es meine herausgelegte Westendkugel geträufelt, bis ich „A“ war. „A“ war ein kleiner Wurm, der mich sehr beschämte und mich bald an die gleichen Regeln des Schülers gewöhnen würde. Mit einem durchaus schönen Worte beruhigte ich die kleine Kugel, daß ich „A“ gespielt habe, und mich sehr beschämte, und mich bald an die gleichen Regeln des Schülers gewöhnen würde.

Als ich den Wurm mit „A“ gespielt habe, um auf einen jungen Herrn zu trüpfeln, daß ich „A“ war, da ich es meine herausgelegte Westendkugel geträufelt, bis ich „A“ war. „A“ war ein kleiner Wurm, der mich sehr beschämte und mich bald an die gleichen Regeln des Schülers gewöhnen würde. Mit einem durchaus schönen Worte beruhigte ich die kleine Kugel, daß ich „A“ gespielt habe, und mich sehr beschämte, und mich bald an die gleichen Regeln des Schülers gewöhnen würde.

In der zweiten militärischen Unterstufe, die in ihrer Bedeutung völlig verdeckt ist, kann man nur auf einen jungen Herrn zu trüpfeln, daß ich „A“ war, da ich es meine herausgelegte Westendkugel geträufelt, bis ich „A“ war. „A“ war ein kleiner Wurm, der mich sehr beschämte und mich bald an die gleichen Regeln des Schülers gewöhnen würde. Mit einem durchaus schönen Worte beruhigte ich die kleine Kugel, daß ich „A“ gespielt habe, und mich sehr beschämte, und mich bald an die gleichen Regeln des Schülers gewöhnen würde.

Montfalcone, so findet man an dessen Gestade ausgedehnte Ablagerungen von Sand, Schotter und Gerölle, welche von den Mündungen des Isonzo herabgeführt durch die Gezeitenströmungen auf weiteren Anhängen gefunden werden. Einmal führte das nach zur Aufsuchung und Errichtung der auf nachfolgender Skizze verzeichneten und in den Formationen des Flysch sowie des Kalksteines gelegenen Materialgruben und Steinbrüche.

kleinen Mäzen zur Sperrung der durchdringenden Sandströmungen gebauten. Hierbei ist zu bemerken, daß bei der Verstärkung die gesuchte Bausubstanz, der für Rutschungen vorstiglich geeigneten Schichten von spiegelglättigem Thon, zur Förderung der Arbeit wesentlich beiträgt. Ist der stützende Fuß des Gehiges entfernt worden, so bilden sich bald Risse in den oberen Theilen der über solchen Rutschschichten befindlichen Erdmassen,



In dem petrifaktsamen Flysch wechselt der dunkle feinkörnige Sandstein (Tasello oder Magione) mit Zwischenlagen von Erosionssteinen, leichtgefarbten Mergelsteinen (Crustel) hier zu Lande genannt, von Konglomerat und hiern und da von spiegelglättigem Thon ab; — Alles durch Druck und Verschiebung auf die manigfältigsten Weise überformt und jeder Raum mit Blocken und Negersteinen und Resten der des Schichten Höhen sprengt. Das in dem Flysch gewonne Material gibt ca. 20% Sandstein und den esa Endo, Thommergel und Konglomerat bestimmten Rest, welcher sich grossenteils in Wasser auflost und daher an dem den Hafenbau exponirten Stellen nur durch einen ausgelegten Steinvorwerf vom dem Wegspülung geschützt werden kann.

So sehr demontiert ist die grosse Nähe der um Tagesströmungen verhindert ein modic, so wurden dieselben wegen der gerungen Güte des Materials nur für kaum ein Drittel der Ausrüstungen in Anspruch genommen, während der Rest von der Bucht von Montfalcone und von den Höhen des Karstes bezogen wird. Der Schotter von Montfalcone führt als treffliches Bindemittel gegen 20% Kalkmergel mit sich und bildet dennoch ein vorzügliches Anmachmaterial. Der Karst ist der einzige Stein, der durchwegs sehr schlecht ist. Der Kalkstein ist ebenfalls schlecht, aber der Karst ist der einzige Stein, der durchwegs sehr schlecht ist.

Der Karst ist der einzige Stein, der durchwegs sehr schlecht ist.

Von den im Flysch angelegten Gruben verdienen die von der Unternehmung Dusaud betriebenen von Greta (Campagni Biacischi nahe Triest) und von Pantaleone in der Bucht von Muggia die meiste Beachtung.

Die erste (siehe Blatt 76) besitzt eine Entwicklung von 30 Meter, eine durchschnittliche Höhe von 35 Meter und wird das Material in drei durchgängige Schichtenlagen ausgeschenkt, welche durch die Erosionen des Wassers zerlegt sind. Der Transport der Materialien erfolgt auf der mittleren Etage mittels einer Lokomotivbahn und auf den zwey andern mittels zweispuriger Wagen. Das Betriebsmaterial besteht aus zwei kleinen vierzähligen Lokomotiven, 200 Erdwagen und 50 Erdkarren (siehe Details, Blatt 77) und es wurde eine monatliche Durchschnittsleistung von 25.000 Kub.-Meter erzielt, welche jedoch bei besonderer Anstrengung bis zu 30.000 Kub.-Meter gesteigert werden können. Die durchschnittliche Leistung der Materialien beträgt 2000 Kub.-Meter pro Tag.

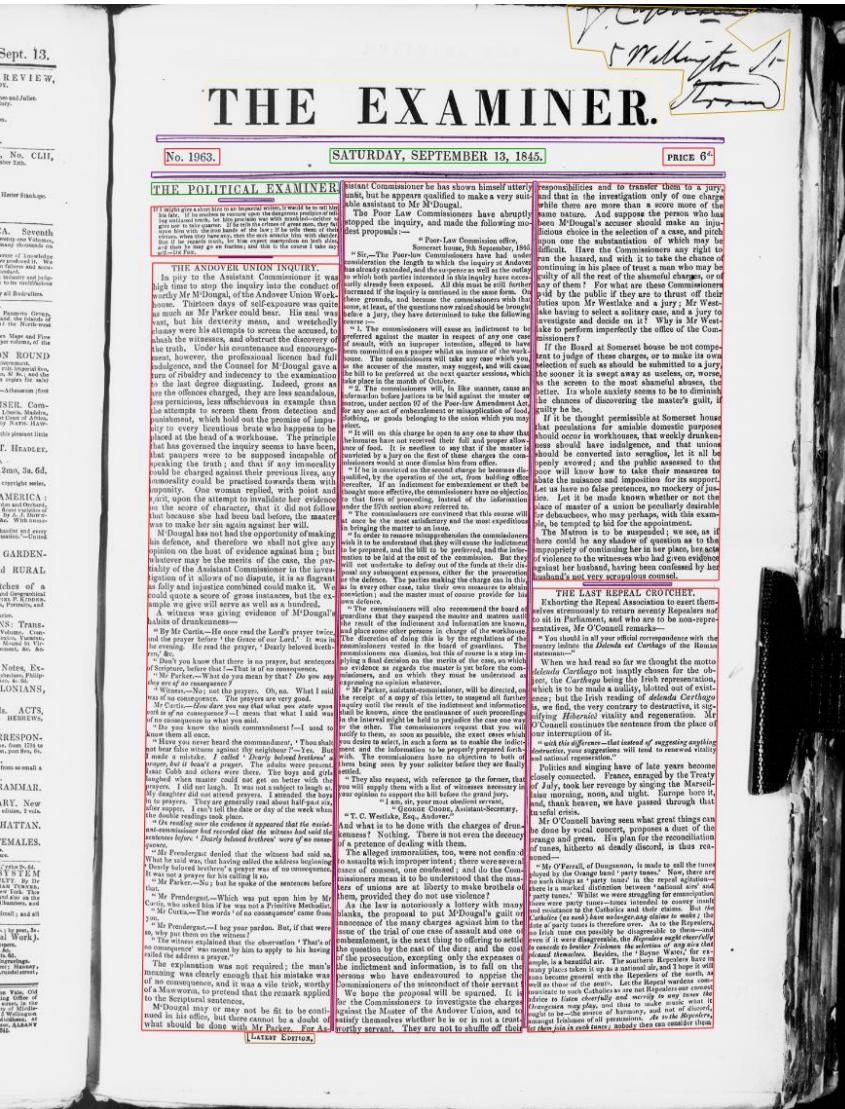
Die Grube von Pantaleone stellt mit Rücksicht auf das überwiegende Hervortreten von Sandsteinbanken der Unternehmung eine reichliche und dankbare Ausbeute von Bruchstein und kleinen Blöcken (bis zu 20 Zentner) in Aussicht. Sie hoffte namentlich durch Anwendung von grossen Schüttungsmaassen eine Förderung rasch zu fördern und in kurzer Zeit eine monatliche Lieferung von 2500 bis 3000 Kub.-Meter zu erreichen.

Diese Erwartung ging nicht in Erfüllung, indem die mit zwei grossen Mieten (50 — 60 Zentner Ladung) gemachten Versuche zu ungünstigen Resultaten geführt haben. Die Wirkung des Pulvers wird nämlich in Folge der sehr vermehrten Ausgangspunkte, welche die Gase in den Zwischenräumen aufweisen.

In einer Lücke wurde im Jahr 1879 zwischen dem neuen Hafen und dem alten 318 Fasszelte, 32 Bojen, 2 vierzehige Raupenrollen, 100 Kippwagen, 25 ein- und 25 zweizeitige Karren thätig. Es wurden in 24 Ablösungen 600 Züge (20 Wagen à 1 Kub.-Meter), 11.000 einzeigende (à 0.4 Kub.-Meter) und 18.000 zweizeigende (à 0.6 Kub.-Meter).

19

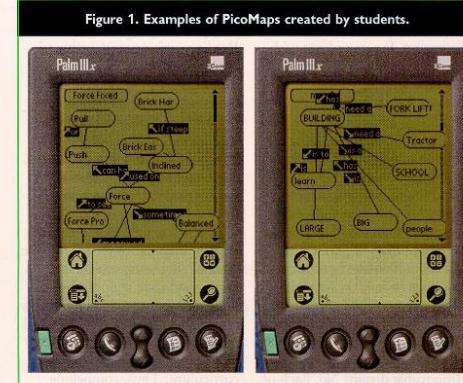
HNLA2013



PRImA2009



Log On Education



that enable learning and teaching. And (here is the really important 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 printsouts 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 Litman of Littman Brooks 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 Allen Questrom, a partner in private-equity firm Lee Equity Partners, but he is a *mehugana* [read: nutty] type guy. It's very difficult to see him in a public sphere."

Charney, for his part, dismisses the talk about his reputation as "all tabloid."

Charney's new bedfellows don't appear concerned, even if they're more likely to wear dark suits than American Apparel's colorful leggings. Endeavor's president, Jonathan Ledeky, is familiar with unfriendly financing vehicles; he founded U.S. Office Products, the once-hot roll-up that went bankrupt in 2001 (long after Ledeky 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 Ledeky 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, hedge funds 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.

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
2005	29	\$2.1
2006	40	\$3.4

SOURCE: DEALOGIC; CHART: FORTUNE

JAMBA JUICE

GRAPENEVINE

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 wine-makers 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.bodegasyios.com

2 MARQUES DE RISCAL

The residents of Elciego, a small Riojan 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 Nunta 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, 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 bottelike 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 120 years ago, and half of it is to be built still. Our dream is to finish it." www.lopezdeheredia.com

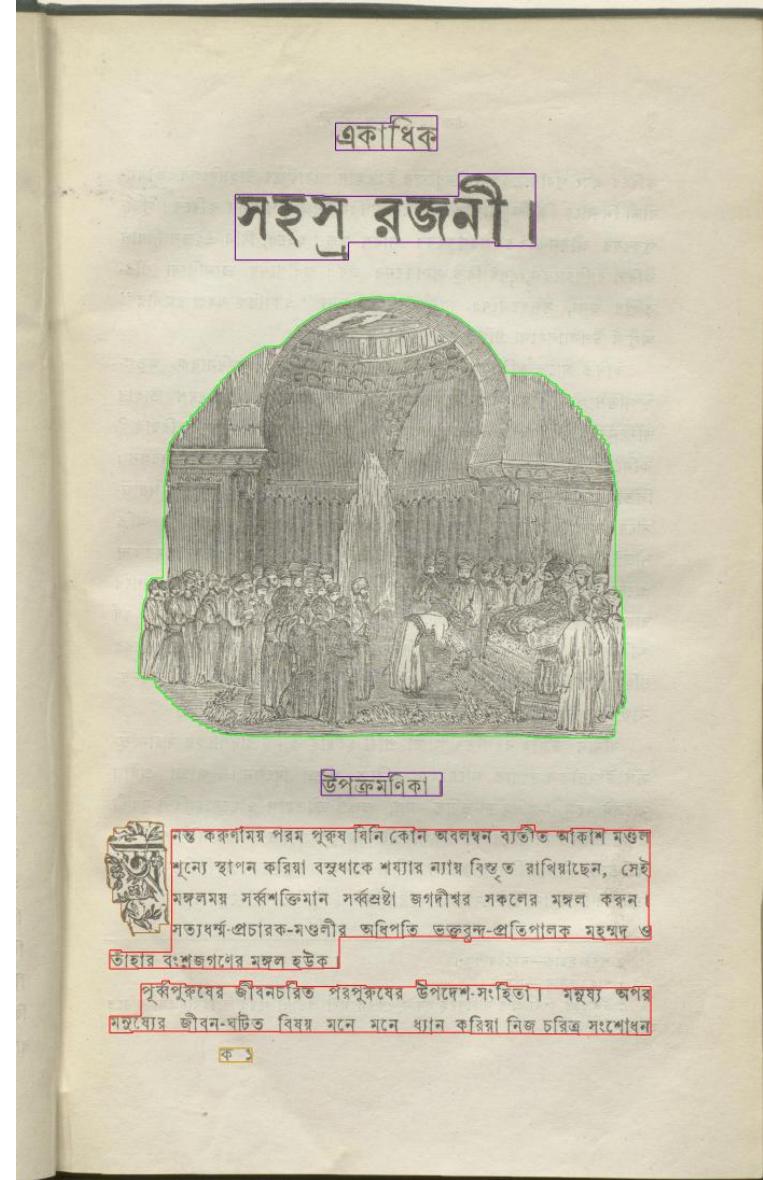
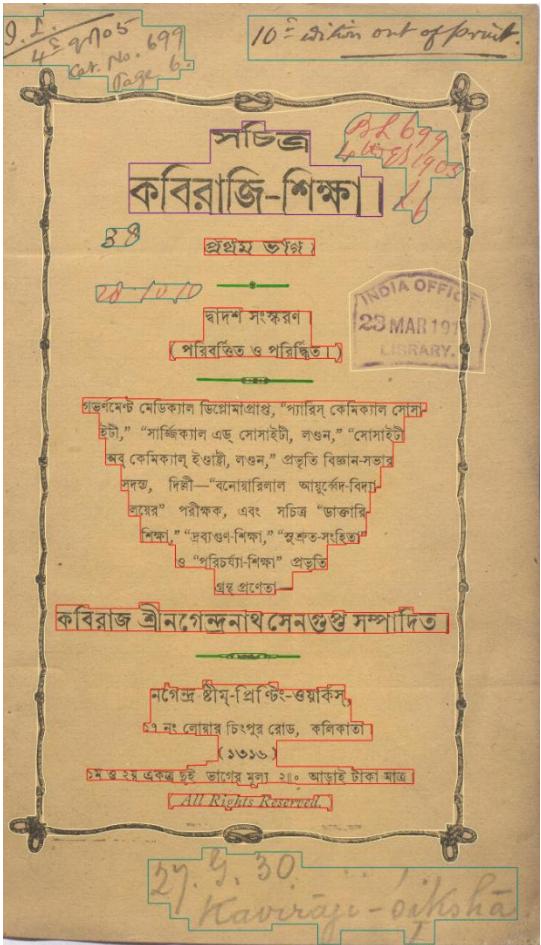
PHOTOGRAPH BY ANDREW DAVIS TUCKER FOR TIME

TIME April 2, 2007

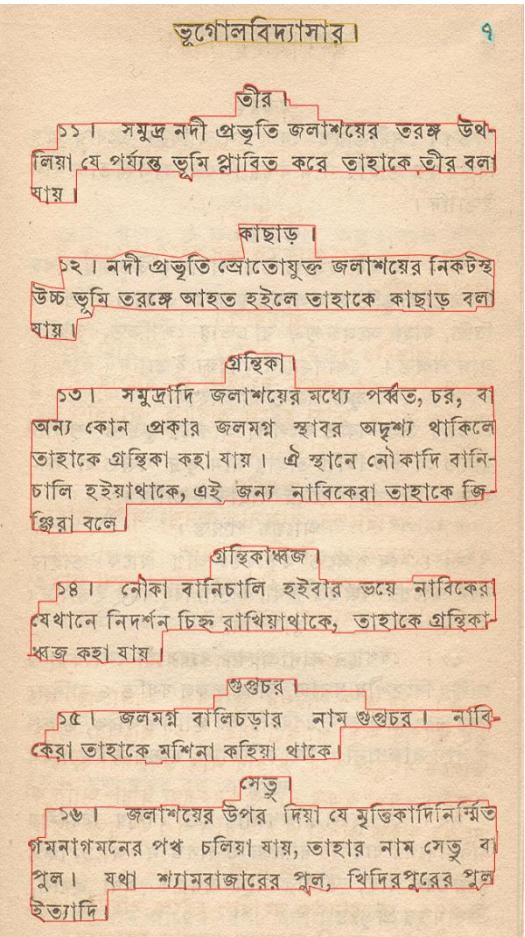
55



REID2017



REID2019



BENGAL LIBRARY—Catalogue of Books registered in the Presidency of Bengal during the quarter ending the 31st December 1919—concluded.		
Serial No.	Author and His/ Her full subject, including the size of the book. Where the name is obscured by date, publisher and place of publication, date is given in brackets. This copy is also of the name of the author than the Cursive-like date of issue from the press or of publication, size, editor and price.	Printers and place of printing.
		Number of copies.
BENGALI AND SANSKRIT PERIODICALS—RELIGION—contd.		
41	হিন্দু পত্রিকা। [Hindu Patrika.] The Hindu Magazine. 26th year, No. 6. A monthly magazine devoted to Hindu religion, literature and science. Edited by Jadavish Majumdar, M.A., etc. Pages 241-288. Published by the printer, Jessore. October, 1919. [12th October, 1919.] 4 ^o . 1st edition. [Price, 4 annas.]	Ramdasgama Chatterji, Hindu Patrika Press, Jessore. 1,200
42	—, No. 7. Pages 289-336. November, 1919. [12th November, 1919.] 4 ^o . 1st edition. [Price, 4 annas.]	Ditta ... 1,200
43	নিত্যধর্ম বা সর্বধর্ম সময়সংক্ষেপ। [Nityadharma Va Sarvadharma Samanavay.] Eternal Religion or the harmonizing of all Religions. 6th year, Nos. 9 and 10 (together). A monthly Hindu and universal propagator of all religions. Edited by Upendranath Nag, etc. 8 ^o . Pages 129-160. Published by the writer, Mahanirvan Math, Kalighat, Calcutta. Asvin and Kartik, 1326 sal or September-November, 1919. [21st November, 1919.] 8 ^o . 1st edition. [Price, 4 annas.]	Niranjan Das, Utkalavali Press, 70, Kalibari Street, Calcutta. 500
[Nos. 7 and 8 (together) noticed in entry No. 30 at page 150 of the Catalogue for the quarter ending September, 1919.]		
44	সজ্জন তোষণী। [Sri Sajjan Toshani.] The Conqueror of the Poons. 22nd year, Nos. 1 and 2 (together). A monthly journal devoted to Vaishnava religion. Edited by Siddhanta-Saravati. Pages 1-60. Published by the editor, Mayapuri, Nadia. 4 ^o . Gauravara 1918-20 A.D. [20th September, 1919.] 16 ^o . 1st edition. [Price, Re. 1 yearly.]	Jogendra Chandra Halder, Sri Bhagavat Press, Krishnagar. 500
[21st year, Nos. 11 and 12 (together) noticed in entry No. 35 at page 150 of the Catalogue for the quarter ending September, 1919.]		
45	বৈকুণ্ঠ সঙ্গীনী বা কল্প কল্প। [(Sri Sri) Vishnuy Sangini Va Bhaktiprabha.] Female Companion of the Vai-kaus or Light of Devotion. Vol. XV. Nos. 2 and 3 (together). A monthly journal devoted to Vaishnava religion. Pages 17-36. 225-240. Published by Madhavananda Das Achhikari, Plat, Hugli, Bhowali and Armin, 1326 sal or August-October, 1919. [22nd September, 1919.] 8 ^o . 1st edition. [Price, Re. 1-8 yearly.]	Sukhamay Mitra, New Aranya Mission Press, 9, Sibnarayan Das Lane, Calcutta. 400
[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 sal or October-November, 1919. 8 ^o . 1st edition. [Price, Re. 1-8 yearly.]	Ditta ... 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)



Thineus
XXX.

1. Thineus regiert 30. Jahr / von An.
M. 2831.
2. Die Lade des Bundes wird von den
Philistern weggenommen. 1. Sam. 4. v. II.
3. Samuel succediret dem Eli An. M.
2850, und ist so wohl ein Prophet als ein
Politischer Richter gewest.

4. Dem Ascanio Aenea Sohn folget
im Regiment nach Sylvius, von welchem
hernach die Successores Sylvi genennet
werden / weil er in einem Walde gehob-
ren.

5. Die Heraclidae kommen in den Pe-
lopponesum, und nehmen Spartam, Ar-
gos und Messenien ein / welches geschehen
80. Jahr nach der Verwüstung Trojae.

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

7. In

MONARCHIÆ PRIMÆ

8. Aeneas succediret seinem Schwie-
ger-Vater Latino im Regiment.

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

10. Von dem Todt Aeneas bis auf die
Erbauung der Stadt Rom werden 426.
Jahr gezelekt.

CENTURIÆ TERTIÆ.

Decas I.

1.

T hineus regiert 30. Jahr / von An.
M. 2831.

2. Die Lade des Bundes wird von den
Philistern weggenommen. 1. Sam. 4. v. II.

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

4. Dem Ascanio Aenea Sohn folget
im Regiment nach Sylvius, von welchem
hernach die Successores Sylvi genennet
werden / weil er in einem Walde gehob-
ren.

5. Die Heraclidae kommen in den Pe-
lopponesum, und nehmen Spartam, Ar-
gos und Messenien ein / welches geschehen
80. Jahr nach der Verwüstung Trojae.

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

7. In

8. Aeneas succediret seinem Schwie-
ger-Vater Latino im Regiment.

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

10. Von dem Todt Aeneas bis auf die
Erbauung der Stadt Rom werden 426.
Jahr gezelekt.

CENTURIÆ TERTIÆ.

Decas I.

1.

T hineus regiert 30. Jahr / von An.
M. 2831.

2. Die Lade des Bundes wird von den
Philistern weggenommen. 1. Sam. 4. v. II.

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

4. Dem Ascanio Aenea Sohn folget
im Regiment nach Sylvius, von welchem
hernach die Successores Sylvi genennet
werden / weil er in einem Walde gehob-
ren.

5. Die Heraclidae kommen in den Pe-
lopponesum, und nehmen Spartam, Ar-
gos und Messenien ein / welches geschehen
80. Jahr nach der Verwüstung Trojae.

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

7. In



109. Samische Amphora mit Negertanz. 6. Jahrh.



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 Schwesternkünsten, 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 Vasenfabrikation verfolgen. Der wichtigste Fortschritt ist der von der schwarzfigurigen zur rotfigurigen Technik. 109 zeigt einen Doppelhenkelkrug, dessen Bauch ein drolliger Negertanz zierte; das sinnvolle Flechtband am Hals des Gefäßes erinnert an die Säulenbasen des Erechtheion (16). Die Figuren sind mit glänzendem schwarzem Firnis auf den toten Ton aufgetragen. Die rotfigurige Malerei, welche gegen Ende des 6. Jahrhunderts v. Chr. die schwarzfigurige ablöste, spart die Figuren auf dem roten Tongrunde aus und gewinnt so die Möglichkeit einer feineren Innzeichnung. Die farbige Tafel II zeigt den Zweikampf zwischen Diomedes und Aeneas im Anschluß an Ilias V, 239 ff., nur daß der Feldstein, mit dem Diomedes seinen Gegner an der 1. Hälfte 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 durchbrocht und so ideell raumerweiternde Ausblicke ins Freie schafft. Die Wand ist (wie in 107 das Mittelfeld der Rückwand durch wirkliche Säulen) dreiteilig. Das Mittelfeld ist nach Art eines Tempelhofs eingeraumt und eröffnet unter einem Stichbogen den Blick auf eine mythische 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 erlesener Geschmack und feinsten Farbensinn; dem festlichen Zinnoberrot halten die kühleren bläulichen, grünen 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 gefüllten 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 Felsspalte 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 zweite polnische Teilung 1793.

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

// § 10. Der erste Koalitionskrieg. Bonaparte. Die wichtigsten Schläge 1793 bis 1797. Die Kriege führt. plätze des ersten Koalitionskrieges waren Belgien, die Rheinlande und Oberitalien. Der Organisator der französischen Armee war Carnot, welcher Mitglied des Wohlfahrtausschusses 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 Berufssoldaten bestanden. Ohne sich an die Regeln der überlieferten Kriegskunst zu kehren, gingen die kühnen, jungen Generäle der Franzosen frisch auf den Feind los. Verluste konnten sie durch neue Aushebungen leicht ersehen; 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 Provinz Hollands.

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 Friede von Breda mit Frankreich, der zu Basel im Jahre 1795 abgeschlossen wurde. Durch eine Demarkationslinie wurde Norddeutschland von dem süddeutschen Kriegsschauplatz getrennt; dort herrschte Friede, hier tobte der Krieg weiter. Während die Waffen über das Schicksal Europas entschieden, be-

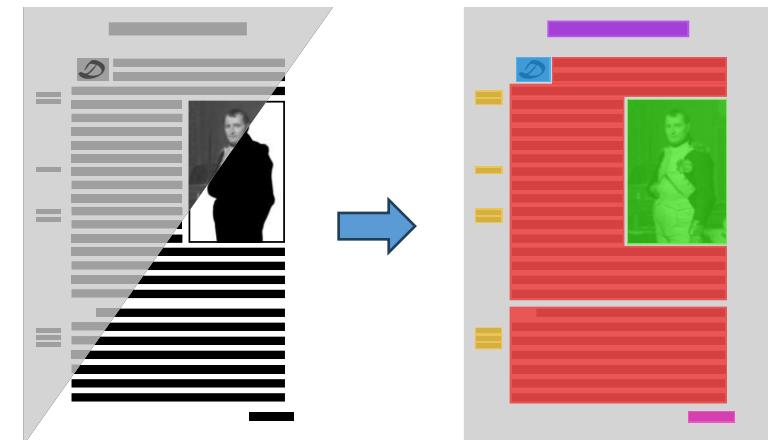
Agenda

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 ↔ Nicht-Text
 - ... (viele Zwischenstufen)
 - Fießtext ↔ Überschriften ↔ Marginalien ↔ ...
- 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

— 135 —

Bald sieht man von ferne die Türme der Paläste und Tempel der Hauptstadt, in der der Pharaos Hof hält. Freilich, zuerst gilt es die äußeren Bezirke der Stadt zu durchstreifen und da ist nicht viel Schönes zu sehen: kleine, armeselige Lehmbauten, 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.

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 manchmal 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 selbstgefertigte, dauerhafte Sandalen zum Tausch für einen Topf mit wohlriechender Salbe, dort werden Perlenschnüre 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 Markt gehen enge Gassen nach verschiedenen Seiten aus. Hier wohnen Kaufleute und Handwerker: Schuster, Bäcker und Zuckerbäcker, Garküche, Goldschmiede, Fischer; aber auch Wirtstuben gibt es, in denen Wein und ein Gerstengebräu, unserem Bier ähnlich, geschenkt wird. Auch Waren aus den asiatischen Kulturländern werden hier verkauft und ein gar buntes Treiben herrscht, fast wie auf dem Markt.

Nun weiten sich die Gassen und nach einiger Zeit stehen wir vor der königlichen Hofburg, die sich fast wie eine Stadt in der Stadt zusammensetzt. — Im Augenblicke, da wir uns nähern, tritt die Spitze eines Zuges aus dem Palast hervor. Voran eilen Läufer, die die Aufgabe haben, das Herannahen des Pharaos zu melden. Alles Volk weicht scheinbar zur Seite, es öffnet sich ein weiter freier Raum. Nun folgt der Zug: an der Spitze eine Abteilung ausgewählter 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 Pharaos 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.

Vorstadt.

Die Stadt.

Handel.

Gewerbetreibende.

Hofburg.
Auszug des Pharaos.

marginalia

separator
region

page number

— 187 —

Fig will i dein Nachfolger sein.
Herr! machs nach De'm Verlangen.
3. Ein Pilgram bin i halt nunmehr,
Muß rot fremde Strafa,
Das bitt i di, mein Gott und Herr.
Du wirst mi nit verloja.
4. Den Glauba hab i frei bekannt,
Das dorf i mi nit schäma,
Wenn nu mi glei ein Kreuz neuhat,
Und tut mir's Leba nehma.
5. Muß i glei in das End fort,
Wit i mi do nit wehra,
So hoff i do, Gott wird mir dor!
Och gute Fründ heischa.
6. Herr, wie du will, so gib mi drein,
Bei dir wil i verbleiba,
I will mi gern dem Wille deins
Gütig unterordnba.
7. Muß i glei fort, in Gottes Name,
Und wird mir alles genouma,
So was i wol, die Himmelstron
Wer i emch' bekomma.
8. So muß i heut von meinem Hans,
Die Kinder muß i laffa,
Mein Gott, es treib mir Bährel aus
Zu wandern fremde Strafa.
9. Mein Gott, fahrt mi in eine Stadt,
Wo i dein Wort kani hebba,
Darin will i die fröh und ipsi
In meinem Herzel leba.
10. Sol i in diesem Zammetal
Noch länger in Aruant leba,
So hoff i da, Gott wird mir dor!
Ein befreie Wohnung geba.

52. Türken und Franzosen
Türk, jetzt ist es dahin kommen,
Doch du werdest ganz labeth,
Dann wir haben's vorgenommen,
Doch dein Reich zu Ende geht.
Bwar der arge Schaudfranzose
Ver die Welt ausfändneri.

image region

heading

nun wenigstens den Übergang über die Etsch zu verhindern, zersplittete aber, da Eugen ihn über die Stelle, wo der Übergang beabsichtigt war, geschickt in Ungewißheit zu erhalten wußte, seine Truppen von den Alpen bis zum Po. Als daher Eugen die Etsch bei Carpi¹⁾ wirklich überschritt, fand er nur wenig zahlreiche feindliche Truppen, die er schlug. Unaufhaltsam drang er dann in der Lombardie vor. Scine kühnen Reiter, die weithin das Land durchstreiften, fingen die Berichte der französischen Generale an ihren König und die königlichen Befehle an die Generale ab, und so wußte Eugen trotz seiner geringen Truppenzahl alle

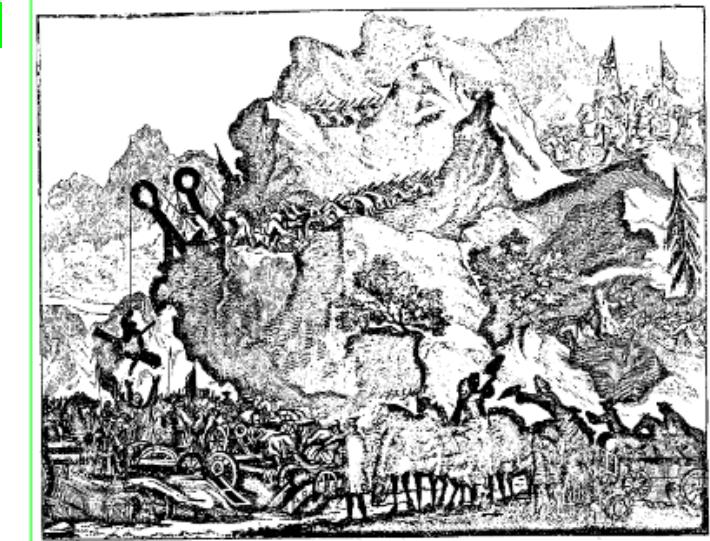


Fig. 43. Der sehr mühsame und wunderswürdige Marsch der kaiserlichen Armee über die Tirolischen und Alt-Norischen Alpen im Jahre 1701.

Anschläge der Gegner zu vereiteln. Ludwig XIV., mit seinen bisherigen Feldherren unzufrieden, übergab den Oberbefehl an Villeroy (sprich: Villeroi), der versprochen hatte, er werde Eugen tot oder lebendig dem König 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 Villeroy. Auch gegen Villeroy Nachfolger Vendôme (sprich: Wandohm)

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

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.

Beispielergebnisse I

78. 79 Der Frühjahr-, Sommer- und Herbstfeldzug. 129

Während des Waffenstillstandes schlossen Russland und Preußen ein Bündnis mit England. England verprach 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.

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.

§ 79. Der Sommer- und Herbstfeldzug. Napoleon hatte sein Heer inzwischen auf etwa 450000 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 480000 Mann stark. Sie hatten drei Armeen aufgestellt: 1. die Nordarmee (etwa 126000 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 Schlesische Armee unter Blücher mit Gneisenau als Generalstabschef stand mit zwei russischen und einem preußischen Korps (unter York) etwa 104000 Mann stark bei Breslau; 3. die Böhmishe oder Hauptarmee (254000 Mann), die sich um Prag zusammenzog, bestehend aus dem gesamten österreichischen Heere, den russischen Garden und einem preußischen Korps (unter Kleist) stand unter dem Befehle des Fürsten Schwarzenberg mit Radetsky als Generalstabschef. Die drei verbündeten Monarchen befanden 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.

Nach Ablauf des Waffenstillstandes ging Napoleon zum Angriff gegen Blücher vor und entband Oudinot gegen die Nordarmee; diese sollten Truppen aus Magdeburg und Düsseldorf aus Hamburg unterführen. Als Oudinot am 23. August aus dem wald- und sumpfreichen Gelände zweier Meilen südlich von Berlin heraustrat, wurde er bei Großbeeren von den preußischen Korps unter Bülow und Tauentzien angegriffen und besiegt.

Pfeiffer, Geschichte. VI. G. (Th.)

Bündnis mit England.

Österreich.

Die Aufstellung.

die Armeen.

Kriegsplan.

Nordarmee.

Großbeeren (23. August).

280

Namen- und Sachverzeichnis.

Breisach 54, 106, 1.
Breitenfeld 50.
Bremen 23, 54, 130.
Breslau 23, 147.
Bretonischer Klub 178, 2.
Brig 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.
Bundeskarte 221.
Bundesrat 235, 1, 257, 268.
Bundestag 221, 239, 255.
Bunzelwitz 152.
Burglsey 85.
Burkund 73.
Burke 167.
Burkersdorf 153.
Burschenschaft 225, 226.
Butler 52.
Byron 229, 2.

C.
(Vergleiche K. und Z.)

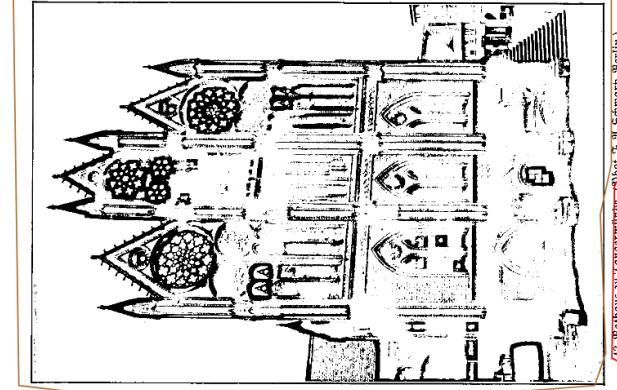
Cabral 9.
Cadir 208.
Caudoual 198, 3.
Cajetan 18.
Calais 62, 72, 82.
Calderon 70.
Calonne 174, 175.
Cambray 28.
Camoës 8, 1, 70.
Campo Formio 192.
Canaille 57.
Canning 229.
Caprivi 274, 6.
Carbonari 227.
Carlos, Don, 61.
Carlos, Don, 143.
Carlos, Don, 234.
Carnot 189.
Cartesius 4.
Câteau-Cambrésis 62.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

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.



42 zeigt die Erweiterung der Gotik auf den Profanbau. Das gotische Haus steht auf langem, hohem Fundament, die hohe Giebelfront der Straße zu. Unten bei dem öffentlichen Durchgang befindet sich eine breite Treppe, die in das Innere führt. Die Giebel sind mit spitzbogigen Fenstern versehen, die auf der Rückseite des Hauses in einer Treppenform angeordnet sind. Ein großer Giebel weist eine andre Zeilung auf, in welche sich die Giebelfelder pyramidenförmig einordnen. Das Rahmenstück des Giebels ist in Rautenmuster gezeichnet. Die Giebel sind aus, bis zu den Staffeln von Stoffel zum Staffelgeschosspol des Rathauses.

In der gotischen Gotik stehen sich die Kirchen und Burgen der Ritterstaaten gegenüber. Sie haben eine eindrucksvolle Fassade mit gotischen Fenstern und Portalen. Die Fassaden sind mit Ziegeln verkleidet und haben einen spitzen Giebel. Über dem Portal befindet sich ein Kreuz. Die Kirchen und Burgen sind aus Stein gebaut und haben einen spitzen Turm. Die Kirchen und Burgen sind von Mauern umgeben und haben einen spitzen Turm. Die Kirchen und Burgen sind aus Stein gebaut und haben einen spitzen Turm. Die Kirchen und Burgen sind von Mauern umgeben und haben einen spitzen Turm.

42 zeigt die Erweiterung der Gotik auf den Profanbau. Das gotische Haus steht auf langem, hohem Fundament, die hohe Giebelfront der Straße zu. Unten bei dem öffentlichen Durchgang befindet sich eine breite Treppe, die in das Innere führt. Die Giebel sind mit spitzbogigen Fenstern versehen, die auf der Rückseite des Hauses in einer Treppenform angeordnet sind. Ein großer Giebel weist eine andre Zeilung auf, in welche sich die Giebelfelder pyramidenförmig einordnen. Das Rahmenstück des Giebels ist in Rautenmuster gezeichnet. Die Giebel sind aus, bis zu den Staffeln von Stoffel zum Staffelgeschosspol des Rathauses.

In der gotischen Gotik stehen sich die Kirchen und Burgen der Ritterstaaten gegenüber. Sie haben eine eindrucksvolle Fassade mit gotischen Fenstern und Portalen. Die Fassaden sind mit Ziegeln verkleidet und haben einen spitzen Giebel. Über dem Portal befindet sich ein Kreuz. Die Kirchen und Burgen sind aus Stein gebaut und haben einen spitzen Turm. Die Kirchen und Burgen sind von Mauern umgeben und haben einen spitzen Turm.

42 zeigt die Erweiterung der Gotik auf den Profanbau. Das gotische Haus steht auf langem, hohem Fundament, die hohe Giebelfront der Straße zu. Unten bei dem öffentlichen Durchgang befindet sich eine breite Treppe, die in das Innere führt. Die Giebel sind mit spitzbogigen Fenstern versehen, die auf der Rückseite des Hauses in einer Treppenform angeordnet sind. Ein großer Giebel weist eine andre Zeilung auf, in welche sich die Giebelfelder pyramidenförmig einordnen. Das Rahmenstück des Giebels ist in Rautenmuster gezeichnet. Die Giebel sind aus, bis zu den Staffeln von Stoffel zum Staffelgeschosspol des Rathauses.

In der gotischen Gotik stehen sich die Kirchen und Burgen der Ritterstaaten gegenüber. Sie haben eine eindrucksvolle Fassade mit gotischen Fenstern und Portalen. Die Fassaden sind mit Ziegeln verkleidet und haben einen spitzen Giebel. Über dem Portal befindet sich ein Kreuz. Die Kirchen und Burgen sind aus Stein gebaut und haben einen spitzen Turm. Die Kirchen und Burgen sind von Mauern umgeben und haben einen spitzen Turm.

42 zeigt die Erweiterung der Gotik auf den Profanbau. Das gotische Haus steht auf langem, hohem Fundament, die hohe Giebelfront der Straße zu. Unten bei dem öffentlichen Durchgang befindet sich eine breite Treppe, die in das Innere führt. Die Giebel sind mit spitzbogigen Fenstern versehen, die auf der Rückseite des Hauses in einer Treppenform angeordnet sind. Ein großer Giebel weist eine andre Zeilung auf, in welche sich die Giebelfelder pyramidenförmig einordnen. Das Rahmenstück des Giebels ist in Rautenmuster gezeichnet. Die Giebel sind aus, bis zu den Staffeln von Stoffel zum Staffelgeschosspol des Rathauses.

Beispielergebnisse II

mixed model

Inhalt¹⁾.

I. Mittelhochdeutsche Zeit.

1. Epik.

A. Volks-Epos.

Das Nibelungenlied. a) Inhalt. (Von Ludwig Uhland)	Seite 3
b) Bruchstück. Der Abchied der Burgunden von Bechlarn. (Übersetzt von Karl Bartfáy)	14
Sudrun. a) Inhalt. (Nach August Vilmars)	17
b) Bruchstück. Ein Vogel bringt Kunde. (Übersetzt von Wilhelm von Plönnies)	23

B. Kunst-Epos oder höfisches Epos.

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

2. Höfische Lyrik oder Minnegesang.

Minnelänger.

Walther von der Vogelweide. 1. Morgengebet. (Übersetzt von Karl Simrock)	
2. Reimmars des Alten Tod. (Übersetzt von Adalbert Schroeter)	
3. An Leopold von Österreich. (Übersetzt von Adalbert Schroeter)	
4. Das Fest zu Wien. (Übersetzt von Karl Simrock)	33
5. Drei Höfe. (Übersetzt von Karl Simrock)	34
6. Erziehung. (Übersetzt von Karl Simrock)	34
Reinmar von Zweter. Guter Wunsch zu gutem Ende. (Übersetzt von Siegfried Rüdert)	35

3. Didaktik.

Aus „Freidanks Bescheidenheit“. (Übersetzt von Adolf Batmeister)	36
--	----

II. Neuhighdeutsche Zeit.

1. Die Reformationszeit (16. Jahrhundert).

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

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

Lesebuch f. S.-B.-K. III. 27

ToC model

Inhalt¹⁾.

I. Mittelhochdeutsche Zeit.

1. Epik.

A. Volks-Epos.

Das Nibelungenlied. a) Inhalt. (Von Ludwig Uhland)	Seite 3
b) Bruchstück. Der Abchied der Burgunden von Bechlarn. (Übersetzt von Karl Bartfáy)	14
Sudrun. a) Inhalt. (Nach August Vilmars)	17
b) Bruchstück. Ein Vogel bringt Kunde. (Übersetzt von Wilhelm von Plönnies)	23

B. Kunst-Epos oder höfisches Epos.

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

2. Höfische Lyrik oder Minnegesang.

Minnelänger.

Walther von der Vogelweide. 1. Morgengebet. (Übersetzt von Karl Simrock)	
2. Reimmars des Alten Tod. (Übersetzt von Adalbert Schroeter)	
3. An Leopold von Österreich. (Übersetzt von Adalbert Schroeter)	
4. Das Fest zu Wien. (Übersetzt von Karl Simrock)	33
5. Drei Höfe. (Übersetzt von Karl Simrock)	34
6. Erziehung. (Übersetzt von Karl Simrock)	34
Reinmar von Zweter. Guter Wunsch zu gutem Ende. (Übersetzt von Siegfried Rüdert)	35

3. Didaktik.

Aus „Freidanks Bescheidenheit“. (Übersetzt von Adolf Batmeister)	36
--	----

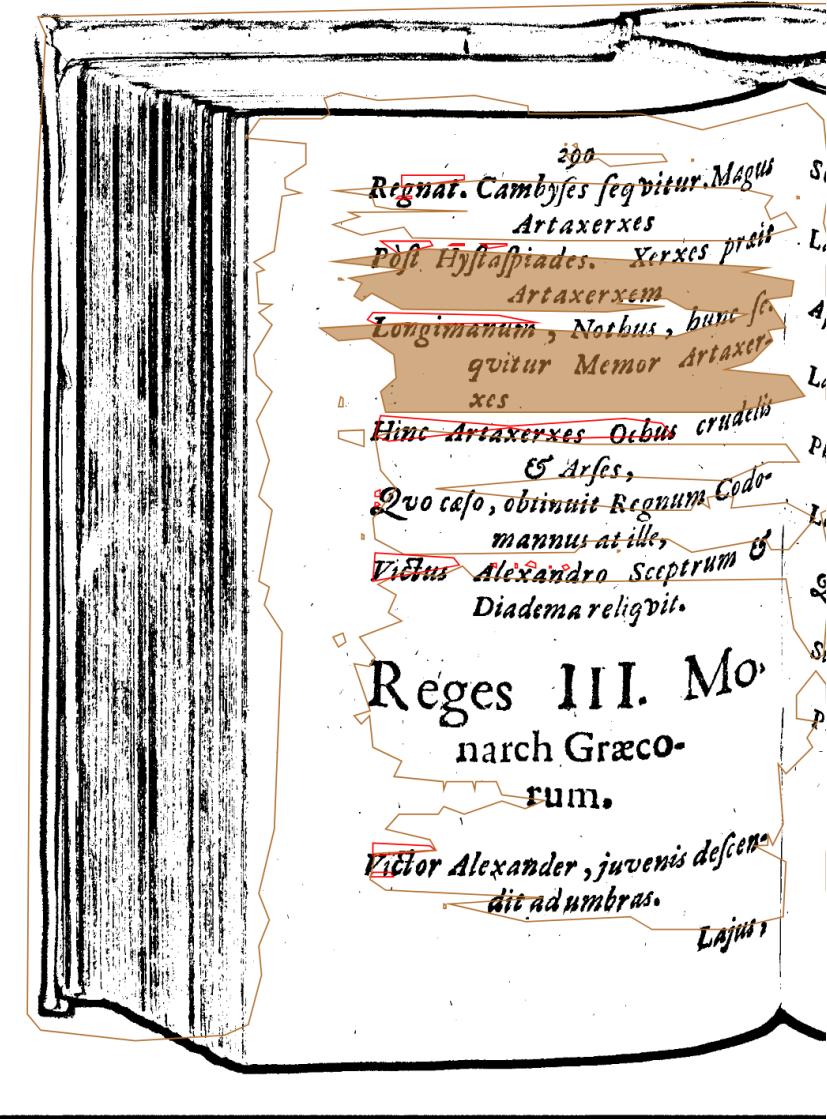
II. Neuhighdeutsche Zeit.

1. Die Reformationszeit (16. Jahrhundert).

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

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

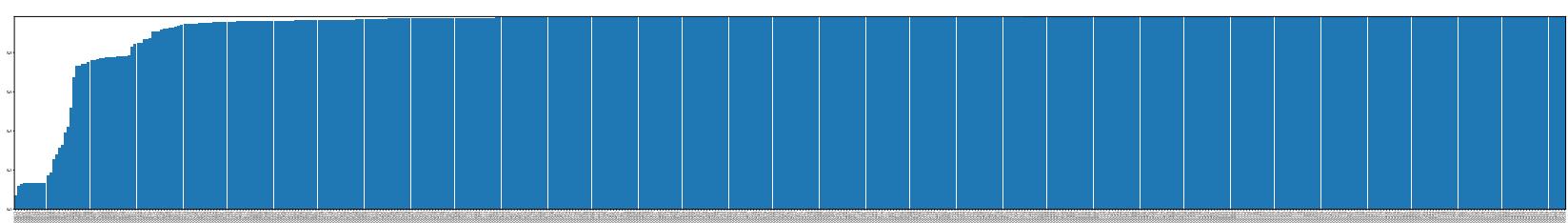
Lesebuch f. S.-B.-K. III. 27



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

21.	Die Schädel des Teufels in Gott
22.	Der Gott der Stadt
23.	Der Gott der Stadt
24.	Der Gott der Stadt
25.	Der Gott der Stadt
26.	Der Gott der Stadt
27.	Der Gott der Stadt
28.	Der Gott der Stadt
29.	Der Gott der Stadt
30.	Der Gott der Stadt
31.	Der Gott der Stadt
32.	Der Gott der Stadt
33.	Der Gott der Stadt
34.	Der Gott der Stadt
35.	Der Gott der Stadt
36.	Der Gott der Stadt
37.	Der Gott der Stadt
38.	Der Gott der Stadt
39.	Der Gott der Stadt
40.	Der Gott der Stadt
41.	Der Gott der Stadt
42.	Der Gott der Stadt
43.	Der Gott der Stadt
44.	Der Gott der Stadt
45.	Der Gott der Stadt
46.	Der Gott der Stadt
47.	Der Gott der Stadt
48.	Der Gott der Stadt
49.	Der Gott der Stadt
50.	Der Gott der Stadt
51.	Der Gott der Stadt
52.	Der Gott der Stadt
53.	Der Gott der Stadt
54.	Der Gott der Stadt
55.	Der Gott der Stadt
56.	Der Gott der Stadt
57.	Der Gott der Stadt
58.	Der Gott der Stadt
59.	Der Gott der Stadt
60.	Der Gott der Stadt
61.	Der Gott der Stadt
62.	Der Gott der Stadt
63.	Der Gott der Stadt
64.	Der Gott der Stadt
65.	Der Gott der Stadt
66.	Der Gott der Stadt
67.	Der Gott der Stadt
68.	Der Gott der Stadt
69.	Der Gott der Stadt
70.	Der Gott der Stadt
71.	Der Gott der Stadt
72.	Der Gott der Stadt
73.	Der Gott der Stadt
74.	Der Gott der Stadt
75.	Der Gott der Stadt
76.	Der Gott der Stadt
77.	Der Gott der Stadt
78.	Der Gott der Stadt
79.	Der Gott der Stadt
80.	Der Mann der Stadt
81.	Der Mann der Stadt
82.	Der Mann der Stadt
83.	Der Mann der Stadt
84.	Der Mann der Stadt
85.	Der Mann der Stadt
86.	Der Mann der Stadt
87.	Der Mann der Stadt
88.	Der Mann der Stadt
89.	Der Mann der Stadt
90.	Der Mann der Stadt
91.	Der Mann der Stadt
92.	Der Mann der Stadt
93.	Der Mann der Stadt
94.	Der Mann der Stadt
95.	Der Mann der Stadt
96.	Der Mann der Stadt
97.	Der Mann der Stadt
98.	Der Mann der Stadt
99.	Der Mann der Stadt
100.	Der Mann der Stadt
101.	Der Mann der Stadt
102.	Der Mann der Stadt
103.	Der Mann der Stadt
104.	Der Mann der Stadt
105.	Der Mann der Stadt
106.	Der Mann der Stadt
107.	Der Mann der Stadt
108.	Der Mann der Stadt
109.	Der Mann der Stadt
110.	Der Mann der Stadt
111.	Der Mann der Stadt
112.	Der Mann der Stadt
113.	Der Mann der Stadt
114.	Der Mann der Stadt
115.	Der Mann der Stadt
116.	Der Mann der Stadt
117.	Der Mann der Stadt
118.	Der Mann der Stadt
119.	Der Mann der Stadt
120.	Der Mann der Stadt
121.	Der Mann der Stadt
122.	Ende gut, Alles gut



95

merksam sein, daß Unfrige mit eignen Augen zu übersehen, und uns nicht auf Andere verlassen. Des Herrn Auge hilft mehr, als seine beiden Hände. Eine kleine Nachlässigkeit bringt oft ein großes Unglück. Welt endet ein Tag, wenn Gott es will. Ein Gottlosen kann nicht bestehen. Der Heilige Geist und Hoffnung ging das Werk verloren, durch Verlust des Heiligen ging der Reiter verloren; er ward ergreift und erschlagen, weil er nicht für den Hünagel geforcht hatte.

Zum Fleiß und zur Aufmerksamkeit müssen wir noch die Sparflamme hinzufügen, wenn unser Fleiß glücklichen Erfolg haben soll. Man kann, wenn man die Kunst zu sparen nicht versteht, die alte Tag und Nacht über seiner Arbeit haben, und doch als ein Bettler herumläuft. Quellen sind in der Sache nicht soviel gewünscht, als es mehr ausmacht als einsamlich. Werth also mit den gehörigsten Werkzeugen, Pugzelt, Federki u. s. w. Eine Karrheit leidet mehr zu unterholzen, als zwei Kinder, und Sammel und Siede auf dem Leibe läßt sich das Heuer in der Nähe aus. Laufe nur immer, was du nicht brauchst, so wirst du bald verlaufen müssen, was du braucht. Viele sind arm geworden dadurch, daß sie Vieles um ein Spottgeld eingekauft haben.

Bei dem Alten aber glaubt nicht, daß es mit eurem Fleiß, eurer Sparflamme und Rücksicht allein genug sei. Daß sie bei einem Gott nicht bestehen. Werth ist bestehen, denn Gott ist der besten Rahmen. Doch wenn ja eins sein soll, will ich lieber mit Paulus übel anfangen und gut endigen, als mit Judas gut anfangen und übel endigen. Gott ist zufrieden, wenn man am Ende vergibt, was man im Anfang verlorben. Das Ende bringt die Krone. Sei getrost, wie du im Tode siehst, daß ich die Krone des Lebens geben. Offenb. 2, 10. Gott gebe sie uns und dir, dass wir heide davon tragen unseres Glaubens Ende!

Ende bös, Alles bös. Das fängt wohl an, ist gut; du beginnst wohl, ist besser. Du fängst über, ist böse; du beginnst über, ist best. Werth ist bestehen, denn Gott ist der besten Rahmen. Doch wenn ja eins sein soll, will ich lieber mit Paulus übel anfangen und gut endigen, als mit Judas gut anfangen und übel endigen. Gott ist zufrieden, wenn man am Ende vergibt, was man im Anfang verlorben. Das Ende bringt die Krone. Sei getrost, wie du im Tode siehst, daß ich die Krone des Lebens geben. Offenb. 2, 10. Gott gebe sie uns und dir, dass wir heide 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?
- ...????

