



Digitalisierung und Erschließung historischer Schulbücher mit Schwerpunkt religiöse Bildung – zwischen Masse und gezielten manuellen Eingriffen

Christian Reul

Forschungsstelle für digitale Anwendungen
in den Geisteswissenschaften (DACHS)

Universität Würzburg

27.11.2024

Agenda

- 1. Allgemeine Projektinfos**
2. Genereller (OCR) Workflow
3. Layoutanalyse
4. OCR-Qualitätssicherung und –Optimierung
5. NER und NEL



Allgemeine Projektinfos

Digitalisierung und Erschließung historischer Schulbücher mit Schwerpunkt religiöse Bildung und Aufbau einer Wissensbasis für die bildungshistorische Forschung

- DFG-Projekt, Förderlinie „Digitalisierung und Erschließung“
- Kooperation mit Dr. Anke Hertling (Leibniz-Institut für Bildungsmedien | GEI)
- Laufzeit drei Jahre, Start 07/24
- Ziele:
 - Erfassung von deutschspr. Religionsschulbücher und schulischen (Erst-)Lesebüchern
 - Knapp 2.000 Bände mit fast 500.000 Seiten
 - Zeitraum 1618 bis 1870
 - Systematische Erfassung von Akteuren aus der historischen Schulbuchproduktion unter Einsatz von Named Entity Recognition (NER) und Named Entity Linking (NEL)

Borrede.

ser, oder einer Geschichte, verschiedene Dinge zu merken, so sind zwar solche in der Ordnung nach einander hergestellt; Dem Lernenden aber kann die Sache erleichtert werden, wann man durch kleinere Fragen Dieselbe annoch zergliedert, und die Antwort solcher gestalt von ihm heraus locket; 3. E. p. 60. werden Caligulae Laster und sonderlich seine Verschwendung beschrieben, weil diese nun, nach einander zu erzählen, einem Anfänger in der Historie schwer fallen möchten, so kann ad §. 1. gefraget werden: Wie viel Geld hat Caligula durchgebracht? ad §. 2. Wie ging er mit seinen Schwestern um? ad §. 3. & 4. Wie tractirte er sein Pferd? und so an andern Orten mehr.

Was die gemengte oder umgekehrte Fragen pag. 22. 29. 36. 45. &c. denen Lernenden für einen Vorteil und Impression geben, wird man bei dem ersten Anblick derselben alsbald urteilen,

MONARCHIÆ PRIMÆ

8. Aeneas succedit regi seinem Schwieger-Vater Latino im Regement.

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

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

CENTURIÆ TERTIÆ. Decas I.

Thineus
XXX.

T h i n e u s r e g i e r e t 30. J a h r / v o n A. M. 2831.

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

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

4. Dem Ascanio Aeneas Sohn folget im Regement nach Sylvius, von welchem hernach die Successores Sylvii genennet werden/weil er in einem Walde geboren.

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

Dercylus
XXXI.

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

7. In

Eroberungspläne zu fördern, und ließ Truppen in Polen einrücken. Um nicht daß ganze Land eine Beute Russlands werden zu lassen, schloß Friedrich Wilhelm II., wie sein großer Vorgänger, mit der Kaiserin einen ^{zweite pol. Teilung:} 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 eifersüchtiger beobachtete es die Vergrößerung Preußens.

Zu diesem entstand in Polen eine starke nationale Bewegung zum Schutze des zerstückelten und in seinem Dasein bedrohten Vaterlandes; an ihre Spitze trat Kosciuszko. Aber trotz verzweifelten Widerstandes erlagen sie dem russischen General Suworow, der Praga, die Vorstadt Warschau, erstürmte und die Hauptstadt darauf einnahm. Nunmehr folgte die dritte Teilung Polens, das aufhörte ein selbständiger Staat ^{dritte Teilung:} zu sein. Preußen erhielt Neostpreußen mit Warschau, Österreich 1793. Westgalizien, Russland die großen Gebiete, die übrig blieben.

§ 10. Der erste Koalitionskrieg. Bonaparte. Die wichtigsten Schauspiäle des ersten Koalitionskrieges waren Belgien, die Rheinlande und Oberitalien. Der Organisator der französischen Armeen war ^{Die Kriegsführung:} 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 Grundfahs der allgemeinen Wehrpflicht zur Geltung, während die Heere der Verbündeten aus Berufssoldaten bestanden. Ohne sich an die Regeln der überlieferten Kriegskunst zu lehnen, gingen die kühnen, jungen Generäle 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 ^{Eroberung Hollands:} 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 Friede von Basel. 1795. Durch eine Demarkationslinie wurde Norddeutschland von dem süddeutschen Kriegsschauplatze geschieden; dort herrschte Friede, hier tobte der Krieg weiter. Während die Waffen über das Schicksal Europas entschieden, be-

326 folgte deshalb ein 22jähriger Krieg gegen die **Sanniter** (326 bis 304, zu der Zeit, als eben **Alexander** das **Persereich** zerstört hatte; vergl. Liv. IX, 17), in welchem die Römer unter manchen Misgeschicken ihre ganze Heldenkraft entwickelten. Nachdem der große sannitische Feldherr **Pontius** ein römisches Heer in den Engpassen von **Caudium** eingesperrt hatte (321), musste dasselbe schimpflich unter dem Joch hingehen; als der **Senat** den **Consul**, welcher den Vertrag eingegangen war, ausliesserte, wies der ehrenhafte **Pontius** diesen zurück. Der altpatrizische **Papirius Cursor** rächte dann die Römer und wurde der Hauptfeldherr dieses Krieges. — Auch die **Etrusker** und **Umbri** hatten sich inzwischen den **Sannitern** angegeschlossen; doch siegte Rom „durch seine centrale Lage und die Festigkeit seines Staatsverbandes“, und im Frieden erkannte **Samnium** die Oberherrschaft der Römer an.

Die Plebejer und Patricier, die unter den schweren Kämpfen inniger verbündet waren, beendeten jetzt ihren Zwist um die Zulassung zu den höheren Amtmännern, indem den Plebejern endlich auch ein gleicher Anteil an dem **Priesterthume** (ogulnischес Gesetz) gewährt wurde (300). Schon während des Krieges war die persönliche **Schuldhaft** aufgehoben. — Ein Versuch des **Appius Claudius Censor** (der sich durch Anlage von **Appia** von Rom nach Capua und der ersten Wasserleitung Volksgut erworben), die nicht grundsätzige Klasse (auch Freigelassene) willkürlich zum Stimmrecht heranzuziehen, um ein patrizisches Regiment auf den Pöbel zu stützen, wurde dadurch unschädlich gemacht, daß **D. Fabius** (hier von **Maximus** genannt) diese neuen Bürger auf die vier städtischen **Tribus** einschränkte, wodurch ihre Stimmen von denen der alten (aerbauenden) Plebejer in den ländlichen **Tribus**, deren Zahl bis 241 allmählich auf 31 stieg, überwogen würden.

d. Dritter (zehnjähriger) **Sanniterkrieg**. Gegen die weitere Ausbreitung der römischen Herrschaft, die immer mehr durch starke Militär-Cölonieen gestützt wurde, traten (298) noch einmal die Völker Italiens zusammen, als sich neue gallische Einwanderer den Etruskern zu Söldnern anboten und diese wie die Umbri sich mit den Sannitern verbündeten. Als die letzteren von Umbrien (im Osten der Apenninen) gegen die Römer ziehen wollten, siegten diese unter **D. Fabius Maximus** bei **Sentinum**, indem sich **Decius Mus** der Sohn dem Tode wöhnte. Nachdem endlich der edle **Pontius** gefangen und schmälerlich hingerichtet war, mussten die Sanniter, Etrusker und Umbri die Oberherrslichkeit der Römer anerkennen. **M. Curius Dentatus**, der den Frieden mit den Sannitern erkämpfte, unterwarf auch noch die **Sabiner**; dann brachte er für die verarmten Plebejer ein Adergesetz in Vorschlag, das große Kämpfe hervorrief. Nach einer nochmaligen Auswanderung der Plebs (auf **Janiculus**) setzte aber der Dictator **Hortensius** nicht nur **Schuldenentlassung** und **Ackervertheilung** durch, sondern nach einer **Nogation** desselben

HISTORIAE UNIVERSALIS. II⁹

Origenes; **Novatus**, novæ heres eos
auctor, lapsis in persecutione veniam
uit, denegavit, cujus sectatores **Catharos**,
i. puros se dixerunt. At, cum Imperi-
atores negligenter rem administra-
rent publicam, **Æmylius Æmylianus**,
Legionum in Pannonia dux, felicissi-
me Scythas repulit, fudit, ea propter à
militibus creatus Imperator, adversus
qvem Gallus & Volusianus cum ex-
ercitu profecti, à militibus suis, spe
majoris præmii ad Æmyliani partes
transiuntibus, occisi sunt; milites ve-
rò trimestri huic Imperatori, qvod de-
derant, Imperium cum vita abstuler-
bus, studio Valeriani. Symbolum.
Galli: Nemo amicus idem & adulator.
Volusiani: Non sine vano publica fama,
non temere est & planè negligendum,
fama publica qvod dictat. **Æmylia-
ni: non gens, sed mens. Virtus scil. non**
sangvis generositas.

28. **Lacinius Valerianus & Gallienus** 25⁸

**F.I.A.C. 253. Ad supremam ille dignita-
tem per omnes honoris gradus ascen-
sione.**

F dit;

Des XVII. SECULI

Der ÖSTERREICHISCHEN KAISER.

Tab.

DES DEUTSCHEN REICHS STÄNDE.

OBERÖSTERRISCHES KREIS. a. Gött. und Komtuarische Städte. Nach Kaiser Karl I. folgt a. den Bändern Johannis Geviatis, der den Salzmarkt Böhmen besaß. Der Sohn Sigismund war II. Sohn, a. d. d'elsten z. Thron. Fürst der Böhmen in Merseburg, der Thüringie erhielt. Böhmen Regenten a. Sohn, der Sohn Job. Georg III. der die Hälfte von Wien verbaute, a. d. d'elste Herzog von Sachsen, 17. Jahrh. brachte auch viele deutsche Kinder Friedens Augustus, Maximilian, Maria, Johanna und Isabella u. Wittelsbach.

32. **MATTHIAS**, sein Sohn, half sich zu Kaiser und König von Böhmen, und der Sohn Johanns Erzherz. 26. und mit dieser Ehre führt a. 27. verschwunden.

33. **Karl Brandenburg**, Johann Albrecht folgt a. k. Karls Ehre an Hofstaat a. 24. Reform, k. 19. Der Sohn Friedrich III. liest die Prachtthüll der Böhmen föhlen, und erlangt a. 29. Dürschnig.

34. **Pommern**, Bogislav folgt a. 17. ohne Ehre da ihm durch den Wettbewerb Böhmen Brandenburg. Hin- tergrund der Böhmen war ein großer Stein.

35. **Albrecht**, Johann Geviatis lebt Domherren, a. 17. Der Sohn Job. Johannes Calixtus hat das erste Etat. Leopoldus a. 32. Der Name erinnert. Johann Georg Weißer aber führt die Linie in Böhmen und Sachsen.

NIEDERÖSTERRISCHES KREIS. a. Herzogtum Brabantwörth und Lüneburg.

Der Sohn Friedrich a. 32. in Württemberg. Danach folgt a. 36. an Kaiser Karl IV. der Sohn Heinrich. Der Sohn Rudolf a. 12. in die Erb. des Böhmen. Die Böhmen erhielten den Pfalzgrafenstand.

V. a. Georgenland, der Sohn, a. 22. Pfalzgräf. von Brandenburg, zur Anfang. Erfurt ist der Sohn a. 16. der Sohn Friedrich. Kaiser Karl V. a. 25. Malchus hat den Kurfürstenstand, a. 26. Pfalzgräf. von Brandenburg, a. 27. der Sohn des Erzherz. Michaelis, mehrheitlich erhielt a. 28. Clausland, a. 29. den Praggräf. Freuden annehmen. Das folgt a. 30. der Sohn des Erzherz. Johann a. 29. Pfalzgräf. von Brandenburg, der Sohn Leopold erhält a. 31. Schloss und Burghausen Augsburg in Westphalen. Leopold II. a. 32. erhält. Der Sohn Georg, Niklas folgt a. 33. Pfalzgräf. von Brandenburg, a. 34. Pfalzgräf. von Brandenburg. Der Sohn a. 35. Pfalzgräf. von Brandenburg.

36. **Ferdinand II.**, a. 36. Pfalzgräf. von Brandenburg, der Sohn a. 37. Schloss und Burghausen a. 38. nach Christian Leopold. Er erhält die Linie aus Böhmen und Sachsen. In Westphalen und Brandenburg.

37. **Wittelsbach**, die Herren werden a. 38. in der Böhmen erlangt. Wallenstein ist a. 39. alte hattet. We-

ber der Krieg in Schlesien erlangt. Paderborn, Minden und Tecklenburg mit der Mitwirkung der Böhmen folgt a. 40. Pfalzgräf. von Brandenburg und Sachsen-Paderborn.

38. **Albertus** folgt a. 41. Paderborn und Sachsen-Paderborn, a. 42. Pfalzgräf. von Brandenburg, a. 43. Herzogtum Westfalen.

39. **Wittelsbach** schafft a. 44. Pfalzgräf. von Brandenburg, a. 45. Herzogtum Westfalen.

40. **Wittelsbach** kommt der Erbthüll der Böhmen und Sachsen-Paderborn.

a. 46. **Wittelsbach** wird a. 47. Pfalzgräf. von Brandenburg, a. 48. Herzogtum Westfalen.

b. 47. **Wittelsbach** kommt der Erbthüll der Böhmen und Sachsen-Paderborn.

c. 48. **Wittelsbach** kommt der Erbthüll der Böhmen und Sachsen-Paderborn.

49. **Wittelsbach** folgt a. 49. Pfalzgräf. von Brandenburg, a. 50. Herzogtum Westfalen.

50. **Wittelsbach** folgt a. 51. Herzogtum Westfalen.

51. **Wittelsbach** folgt a. 52. Herzogtum Westfalen.

52. **Wittelsbach** folgt a. 53. Herzogtum Westfalen.

53. **Wittelsbach** folgt a. 54. Herzogtum Westfalen.

54. **Wittelsbach** folgt a. 55. Herzogtum Westfalen.

55. **Wittelsbach** folgt a. 56. Herzogtum Westfalen.

56. **Wittelsbach** folgt a. 57. Herzogtum Westfalen.

57. **Wittelsbach** folgt a. 58. Herzogtum Westfalen.

58. **Wittelsbach** folgt a. 59. Herzogtum Westfalen.

59. **Wittelsbach** folgt a. 60. Herzogtum Westfalen.

60. **Wittelsbach** folgt a. 61. Herzogtum Westfalen.

61. **Wittelsbach** folgt a. 62. Herzogtum Westfalen.

Anmerkungen.

Der Westphälische Kreis besteht a. 48. der Aachener Linie und der Arnsberg Linie und der Marien Linie. Er hat hier drei Hauptlinien.

a. 49. machen die Provinzen mit Spanien einen zwischen Südwand, Provinz Spanien a. 50. die Provinz mit den Provinzen mit Spanien.

b. 50. der Sohn ist der Sohn der Provinz mit Spanien.

c. 51. kommt der Sohn der Provinz mit Spanien.

Agenda

1. Allgemeine Projektinfos
- 2. Genereller (OCR) Workflow**
3. Layoutanalyse
4. OCR-Qualitätssicherung und –Optimierung
5. NER und NEL



Genereller (OCR) Workflow

- Binarisierung/Preprocessing mit SBB-Binarize und OCropus
- Layoutanalyse → nächster Abschnitt
- Erkennung voraussichtlich mit Calamari, vorab allerdings noch Tests; Neue gemischte Modelle werden trainiert/evaluiert
- OCR-Qualitätssicherung und -Optimierung → übernächster Abschnitt
- NER und NEL → überübernächster Abschnitt

Agenda

1. Allgemeine Projektinfos
2. Genereller (OCR) Workflow
- 3. Layoutanalyse**
4. OCR-Qualitätssicherung und –Optimierung
5. NER und NEL



Überblick

- Ursprünglicher Plan: OCR-D Workflow Suche
- Aber: lieber direkt voll auf trainierbare Methoden setzen
 - Besser und nachhaltiger
 - Paralleles [Projekt zur Layoutanalyse](#) (mit Stabi Berlin und SLUB Dresden) → Synergieeffekte
 - aktuell Fokus auf Kraken, später ggf. Eynollah u. a.
 - gezielt Trainingsdaten erstellen, Active Learning
- Aktuell viel ausprobieren, Erfahrungen sammeln
- Erste Experimente
 - Pragmatische Regionentypen
 - Recht differenzierte Auszeichnung
 - Für Training vorerst stark eingedampft (Graphic Region, paragraph, marginalia)
 - Roher Output, kein zusätzliches Postprocessing



Beispiele – Trainingsdaten I

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 andere 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 Gerstenbier, 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.

marginalia

Gewerbetreibende.

Hofburg.
Auszug des Pharaos.

Separator
Region

page number

— 187 —

52. Türken und Franzosen
Biegler: Deutsche Soldaten-Kriegslieder aus 5 Jahrhunderten (1380–1871) Leipzig, 1884.
Dr. M. Schilling: Liederbuch, S. 219/20

Türk, jetzt ist es dahin kommen,
Doch du werdest ganz labeth,
Dann wir haben's vorgenommen,
Doch dein Reich zu Ende geht,
Swarz der arge Schandfranzose,
Der die Welt ausfänderviert,

Igt will i dein Nachfolger sein,
Herr, machs nach De'm Verlanga,
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 hob i frei bekannt,
Das dorf i mi nit schâma,
Wenn mu mi glei ein Kreuz neuat,
Und tut mir's Leba nehma,
5. Muß i glei in das Endj jor,
Wit i mi do nit wehra,
So hoff i do, Gott wird mir dor
Och gute Fründ heishera,
6. Herr, wie du will, so gib mi dreim,
Bei dir wil i verbleiba,
I will mi gern dem Wille deim
G'duldig unterzireiba,
7. Muß i glei fort, in Gottes Name,
Und wird mir alles genouma,
So was i wol, die Himmelstron
Wer i emchj bekomma,
8. So muß i heut von meinem Hans,
Die Kinderl muß i laffa,
Mein Gott, es treib mir Bâbel aus
Zu wandern fremde Strafa,
9. Mein Gott, fahrt mi in eine Stadt,
Wo i dein Wort kani heba,
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,

heading

other

135

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

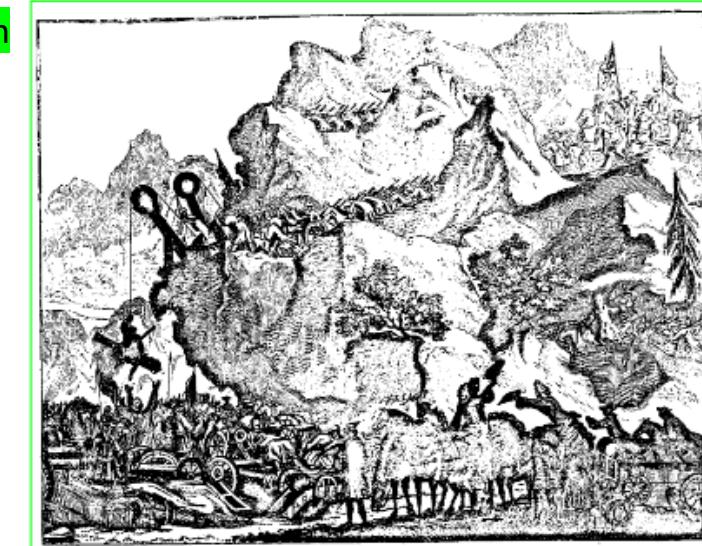


Fig. 43. Der sehr mühsame und wunderswürdige Marsch der kaiserlichen Armee über die Tirolischen und Alt-Norischen Alpen im Jahre 1701.

caption

Anschläge der Gegner zu vereiteln. Ludwig XIV., mit seinen bisherigen Feldherren unzufrieden, übergab den Oberbefehl an Villeroy (sprich: Villero), 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.

footnote

Beispiele – Trainingsdaten II

— 16 —

Rückzug der wichen zurück und ließen sich in keine offene Schlacht ein, sondern Skythen verwüsteten nur das Land und verschütteten die Brunnen, bis den Persern die Lebensmittel ausgingen und Darius den Befehl zum

Rückzug gab. Jetzt begannen die Skythen die Verfolgung, um schwärzten das Heer von allen Seiten und bedrängten es so sehr,

dass es froh sein mußte, wieder den Ister zu erreichen. Zum Glück fand man dort die Brücken noch vor; der Mann, dem Darius seine Rettung verdankte, war Histäus, der Stadtherrsscher (Tyrann) von Milet, einer der griechischen Städte an der Westküste Kleinasiens. Es waren nämlich Skythen an den Fluß gekommen und hatten den Griechen geraten, die Brücken abzubrechen, dann würde das ganze persische Heer zugrunde gehen; Miltiades, ein Griechenfürst vom Hellespont, hatte dem beigeistert, da dann die Griechen Kleinasiens sich vom persischen Joch befreien könnten, aber Histäus widersprach und meinte ungekehrt, die Stadtherrsscher hätten alle Ursache, die Rettung des persischen Königs und Heeres zu wünschen, da sie nur durch des Königs Macht in ihrer Stellung geschützt würden. Ohne diesen Schutz hätten die Bürgerschaften ihrer Städte sie schon längst vertrieben. So blieben die Brücken erhalten und Darius war gerettet. Histäus wurde belohnt, Miltiades floh in seine griechische Heimat.



Fig. 8. Darius mit Dienern (über ihm Ormuz)
Wandbild im Palast zu Persepolis

135

nun wenigstens den Übergang über die Etsch zu verhindern, zersplittete aber, da Eugen ihn über die Stelle, wo der Übergang beabsichtigt war, geschickt in Ungeißheit 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 erschlug. Unaufhaltsam drang er dann in der Lombardei vor. Seine 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

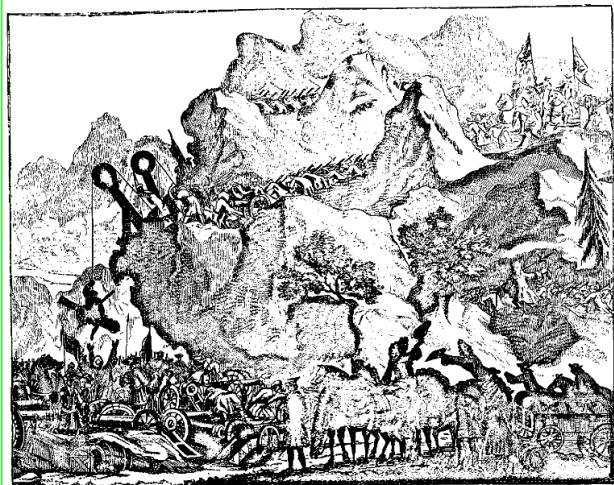


Fig. 43. Der „sehr mühsame und wunderswürdige Marsch“ der kaiserlichen Armee über die Tirolischen und Alk-Norischen Alpen im Jahre 1701

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 Gefangen 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.

— 73 —

nahe stand, doch von ihnen stets als ein barbarisches angesehen wurde. Die wenigen Städte waren meist griechische Kolonien, die Handel und Verkehr im Lande beherrschten.

Philip II. In den Perserkriegen hatten die Mazedonier könige öfters die Vermittler zwischen den Persern, deren Oberhoheit sie damals anerkannten, und den Griechen gespielt; sonst waren sie wenig hervorgetreten. Pelopidas hatte, wie oben erwähnt, Thebens Einfluß auch bis nach Mazedonien ausgedehnt und den jungen Königsohn Philippus als Geisel nach Theben mitgebracht; nach mehrjährigem Aufenthalt daselbst kehrte der Prinz in seine Heimat zurück und gelangte im Jahre 359 zur Regierung.

Philip hatte die Zeit in der Fremde gut benutzt: er kannte die Griechen und ihre Schwächen sehr genau und faßte den kühnen Plan, sich nach und nach der nächstgelegenen Kolonien zu bemächtigen, seine Macht allmählich anzudehnen und dann mit einem Schlag ganz Hellas unter seine Oberhoheit zu bringen. Er war wohl der Mann dazu, das alles auszuführen: ebenso tapfer und entschlossen als gewandt und schlau, scheute er kein Mittel, um sein Ziel zu erreichen; in allen Griechenstädten gewann er sich Freunde, großenteils durch Bestechung; denn, wie er sagte, „einem mit Gold beladenen Esel widersteht keine Festung“. Aus seinen Mazedoniern schuf er sich ein ausgezeichnetes Heer; der Adel, „die Freunde des Königs“, bildete die gut bewaffnete Reiterei und den Kern des Fußvolks, aber auch die freien Bauern dienten eine Zeitlang im Heere als Hopliten oder leichter bewaffnete Fußsoldaten. Sie stellten sich im Kampfe dichtgeschlossen auf, Schild an Schild gedrängt, 16 Glieder tief; die ungewöhnlich langen Lansen (6 m!) ermöglichten es, daß die fünf vordersten Glieder wirklich zum Kampfe kamen, indes die andern nachdrängten; diese mazedonische „Phalanx“ (Schlachtreihe) galt durch lange Jahre als unüberwindlich.

Philip und Athen. Bald hatte Philip unter verschiedenen Vorwänden die bedeutendsten griechischen Kolonien an den Grenzen Mazedoniens eingenommen, ohne ernstlichen Widerstand zu finden. Die Athener, die bisher in diesen Gebieten den Haupteinfluß gehabt hatten, ließen ihn ruhig gewähren; die Stadt war längst nicht mehr, was sie einst gewesen. Die Bürger hatten sich allmählich entwöhnt vom Kriegsdienst, den sie gemieteten Söldnern und Feldhauptleuten überlassen. Wie in jener Zeit überhaupt viele

Philip im
Theben.
Seine Thron-
besteigung.
359 v. Chr.

Philipp's
Pläne.

Sein
Charakter.

Das Heer.

Athens
Schwäche.

Soldner-
wesen.

Beispiele – Ergebnisse I

128 **Das Mittelalter.**

Keine Straßenbeleuchtung. Abends und in der Nacht, wo es keine Laternen gab, und jeder, der das Haus verließ, ein kleines Licht bei sich führte, aufrecht zu erhalten.

Abwechslung. Draußen auf dem Lande und hoch oben in den Delfennestern verließ das Leben im ganzen sehr eintönig. Der Burgherr vertrieb sich die freie Zeit gern mit der Jagd und Begegnungen. Mancher zog auch hinaus in die Welt, um für seine „Herrin“ Abenteuer zu bestehen. Dazu kamen die Hoffahrt und die Turniere. Die „Töpfe“ oder „Dörpchen“ (Dörfer, Bauern) erfreuten sich am Tanz unter der Linde; in den Städten aber entstanden die Feste der wechselseitigen Bürgerlichkeit, die Schützenfeste. Das alles geschah zur hochgelobten Sommerszeit oder im Frühling. Zu den langen, einsamen Winternächten aber, die, weil man den kalten Lust wegen gewöhnt war, die Fenster aufstöpselte, auch den kurzen Tag mit in sich schlossen, war es ein Labsal für jung und alt, wenn Fremde kamen, die etwas zu erzählen wußten, über Spieße, die außerordentlich lange vorgetragen und alte Heldenlieder zu singen und zu sagen verstanden. Diese waren „Fahrende“, die man wohl gering schätzte, die aber niemand missen möchte. Reisen war, falls nicht ein kräftiger König den Landfrieden aufrecht erhält, eine gefährliche und bei der schlechten Verhältnisheit der Wege sehr unangenehme Sache. Aus letzterem Grunde reiste man zu Pferde.

Die Künste. In Nordfrankreich war die sog. gotische Bauweise aufgekommen. In ihr, die in leicht aufstrebenden, der Lust und dem Ruhm geöffneten Bauten wohl die Art der germanischen „Lauben“ nachahmte, wurden herrliche Kirchen, dazu Rathäuser und Kaufhäuser und Kaufhallen errichtet. — Die Eigentümlichkeit der gotischen Bauweise war der Spitzbogen. Da nun spitzbogige (aus Kreissegmenten gebildete) Gewölbe einen geringeren Schub und Druck als die in Halbkreisbögen konstruierten romanischen Gewölbe ausüben, bedürfen sie nur einer Unterstützung der Endpunkte. Und zwar trugen den senkrechten Druck des Gewölbes Pfeiler, den schwachen Seitenhub hielten die den inneren Pfeilern entsprechenden, an der Außenseite angebrachten starken Mauerpfeiler (Strebepfeiler) auf, die, in der Höhe abgestuft, mit einem Giebel oder Türmchen gekrönt waren. Damit war der gebogene Mauerbau aufgelöst in ein reiches, kühn aufstrebendes Baugewirke. Die den Raum abschließenden Wände verwandelten sich in leichte Füllungen, und dieser Wandverschluß wurde durch hohe und weite Fenster durchbrochen.

Für die Glieder und die Schmückung galt der Grundsatz, daß diese in enger Beziehung zum Aufbau stehen, jene das Wesen desselben wieder spiegelnd und in unendlichen Abänderungen zur Darstellung bringen sollten. Die schlanken Pfeiler wurden zu Bündelpfeilern zusammengefaßt und reich gegliedert, sie bestanden aus einem Kern und Säulchen, „Dienste“ genannt, die um ihn herumgestellt waren. Das Kapital zeigte die Form des Kelches; zum Schmuck diente naturgetreuer gebilbtes Laubwerk. Das Kreuzgewölbe erschien als neß- oder sternförmiges Gewölbe. Die Fenster waren zu ausgedehnt, darum teilte man sie in Gruppen; den Raum unter dem Bogenfriesel¹⁾ füllte man mit geometrischen Gebilden, „Maßwerk“,

1) In der späteren Zeit der Gotik bildete man den Bogen in der Gestalt eines Kreuzrundens. Auch gab man da den Seitenschiffen die Höhe der Mittelschiffe.

§ 48. Calvin. Der Jesuitenorden. Die Kirchenversammlung von Trient. 195

„Konvikt“, in denen der Nachwuchs für die Prediger und Lehrer herangezogen wurde. Prächtige Schlösser¹⁾ verkündeten das kraftvolle Selbstgefühl und den steigenden Reichtum der Fürstengeschlechter.

6. Karls Abdankung und Tod (1558). So war der stolze Plan, den der junge Kaiser Karl entworfen, gescheitert; weder eine habsburgische Universalmonarchie noch die dazu unerlässliche Vereinigung der Religion war trotz alles noch so eifrigem und geschickten Wirkens durchzusetzen gewesen. Verbittert über das Fehlschlagen seines Lebenswerkes, zog sich der Kaiser, überdies krank und früh gealtert, vom Schauspiel zurück. Nachdem er bereits im Jahre 1554 seinem Sohne Philipp Mailand und Neapel übertragen hatte, legte er in feierlicher Versammlung der burgundischen Stände auch die Regierung der Niederlande zu jenes Günsten nieder (1555). Im nächsten Jahre verzichtete er auch auf Deutschland und Spanien, von denen jenes Ferdinand, dieses Philipp erhielt (1556). Ohn Mönch zu werden, weilete er vor da ab in der Nähe des Klosters des heiligen Hieronymus bei San Justo im spanischen Estremadura, wo er im Jahre 1558 starb.

Gegen Verpflichtung auf den Augsburger Religionsfrieden wurde in demselben Jahre (1558) Ferdinand zum römischen Kaiser gewählt.

B. Die Gegenreformation. Die Ausbildung der Fürstennacht.

§ 48. Calvin. Der Jesuitenorden. Die Kirchenversammlung von Trient.

Die deutschen Protestanten hätten alle Ursache gehabt, zusammenzuhalten und die unter ihnen bestehenden Meinungsverschiedenheiten nicht bis zur Spaltung oder gar zu politischem Gegenseitig ausscheiden zu lassen. Statt dessen aber verbrauchten sie ihre Kraft in ihrem Haider, den auch eine Fürstenversammlung zu Naumburg a. d. Saale (1561) nicht zu bannen vermochte. Dies war um so verhängnisvoller, als bereits die alte Kirche alle ihre Kräfte zusammenfaßte, um nicht nur dem ferneren Ausbreiten der evangelischen Lehre zähmen Widerstand entgegenzusetzen, sondern auch den Kampf um die Rückgewinnung des verlorenen Besitzes entschlossen aufzunehmen. Die Verteidigung und den Gegenangriff gegen die anhende „Gegenreformation“ übernahmen auf protestantischer Seite die Reformierten katholischer Richtung, deren Hochburg die Stadt Genf war.

1. Calvin. Als es offenbar wurde, daß Savoyen darauf ausgehe, die fast ganz selbständige Stadt Genf sich einzuerleben, bildete sich hier eine Freiheitspartei, die Eidgenoss oder Huguenots²⁾, welche den Anschluß an die Schweiz bewirkte.

1) So in Dresden, München, Stuttgart, Heidelberg. Hier wurde der Ottoneumsbau in den Jahren 1556–1563, der Friedrichsbau 1601–1607 errichtet.
2) Besançon Hugues war ihr Befämpfer.

134 § 47. Renaissance der Wissenschaften und der Humanismus.

2. Der Humanismus in Deutschland. Während in Italien die Renaissance in erster Linie eine gesleigerte Pflege des Schönen und eine verfeinerte Lebensführung im Gefolge hatte, in ihrer Entartung aber auch den geselligen Deckmantel leichsfertiger Sitten bot, gedieb sie in Deutschland zur strengen Schule der Geistesbildung und der Erziehung. Die Humanisten, wie man die Vertreter der neuen Bildungsrichtung nannte, erkämpften hier bald die Lehrstühle der Schulen und Universitäten. Auch neue Hochschulen erstanden in rascher Folge, so in Ingolstadt 1472 (gegründet vom bayerischen Herzog Ludwig dem Reichen von Landshut), in Trier 1473, in Mainz und in Tübingen 1476, in Wittenberg 1502, in Frankfurt a. d. Oder 1506.

Unter den älteren der deutschen Humanisten — sie gaben sich zumeist latinierte oder gracierte Namen — ragen hervor: Rudolf Agricola in Heidelberg, † 1485), Konrad Celtis (in Ingolstadt und Wien, † 1508), Johann Reuchlin (in Tübingen und Ingolstadt, † 1522), Desiderius Erasmus (in Basel, † 1536); außerdem Konrad Peutinger aus Augsburg († 1547), Willibald Pirckheimer zu Nürnberg († 1530) und seine Schwester Charitas, Äbtissin eines dortigen Klosters. Andere Humanisten, wie Ulrich von Hutten († 1523 in der Schweiz) und Philipp Melanchthon († 1560 in Wittenberg), gehören als Streitgenossen Luthers der Reformationsepoke an.

§ 48. Renaissance der Künste.

I. Baukunst.

1. In Italien. Durch die neue Zeitströmung begünstigt, gelangte seit der Mitte des 15. Jahrhunderts zunächst in Italien wieder die Architektur des korinthisch-römischen Baustils zur vorherrschenden Geltung. Im Gegensatz

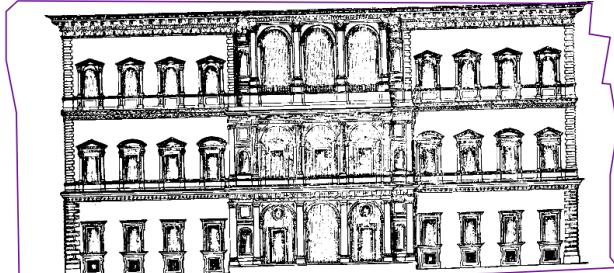


Fig. 11. Der Palazzo Farnese in Rom (Renaissancestil).

zur Gotik schufen die hochgebildeten Meister der italienischen Renaissance lichtreiche Bauten mit ruhigen und harmonischen Grund- und Zierformen.

Beispiele – Ergebnisse II



32. Statue
vom Dom zu Bamberg.

Zu den reifsten Schöpfungen der deutschen romanischen Plastik gehören die im Westchor des Naumburger Domes um 1270 aufgestellten Statuen von 12 fürstlichen Stiftern und Stifterinnen. Eine von diesen stellt 30 dar. Es ist eine Fürstin in Wittentracht, in einem Buche blätternd. Dies Grundmotiv ist für die Gejtanlage der Figur entscheidend. Die r. Hand hat, indem sie das Buch hält, zugleich den Mantel angezogen, so daß er auf dieser Seite in freien großen Faltenzügen das Buch selbst umschreibt und so das Motiv nach unten ausflingen läßt, in scharfem Kontrast zu den senkrechten Falten ihrer l. Seite, die tief herabfallend sich der bereits gotischen Säule angleichen. Das Antik ist durchaus realistisch gehalten, doch in Porträtahnlichkeit durch die Umstände ausgeschlossen. Idealisiert dagegen erscheint das gefrönte Haupt (2 Säcken) das abgebrochen einer Reiterstatue am Dom zu Bamberg, in dem man Konrad III. vermutet (32). Der von einem schlanken Hals getragene, etwas rechts und aufwärts gewendete, von wallendem Haar umrahmte Kopf ist von außerordentlicher Kraft und Frische der Ausführung. — Unruhig wirkt gegenüber der Naumburger Statue das Doppelgräbn Heinrichs des Löwen und seiner Gemahlin Matilde im Braunschweiger Dom (31; f. 20). Die Figuren sind, obwohl sie liegen, doch wie stehend gehoben; andererseits aber fallen die Gewänder auch wieder nach dem Geize der Schwere; es erklärt sich die etwas ängstliche Faltengebung besonders des Frauengewandes. Man erkennt hieraus, daß der Künstler gewohnt und geübt war, Standbilder wie etwa 30 zu schaffen; hier sah er sich nun vor eine ganz neue Aufgabe gestellt, die er nicht widerprüflos zu lösen vermachte. Porträtahnlichkeit ist kaum anzunehmen. Heinrich trägt, wie üblich, in der Rechten das Modell seiner Stiftung, des Braunschweiger Domes.

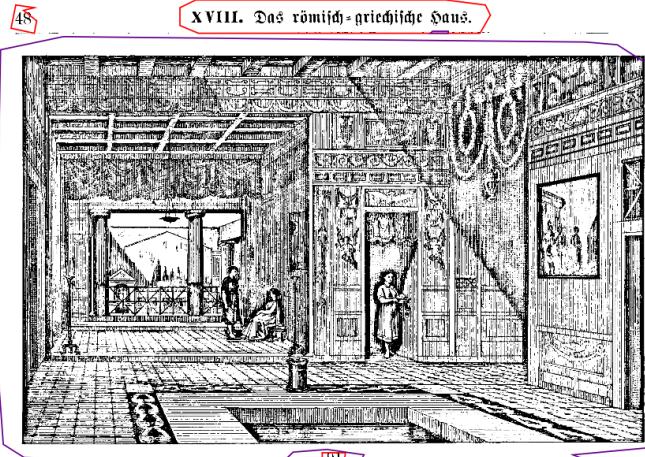
— 15 —



32. Statue
vom Dom zu Bamberg.

Zu den reifsten Schöpfungen der deutschen romanischen Plastik gehören die im Westchor des Naumburger Domes um 1270 aufgestellten Statuen von 12 fürstlichen Stiftern und Stifterinnen. Eine von diesen stellt 30 dar. Es ist eine Fürstin in Wittentracht, in einem Buche blätternd. Dies Grundmotiv ist für die Gejtanlage der Figur entscheidend. Die r. Hand hat, indem sie das Buch hält, zugleich den Mantel angezogen, so daß er auf dieser Seite in freien großen Faltenzügen das Buch selbst umschreibt und so das Motiv nach unten ausflingen läßt, in scharfem Kontrast zu den senkrechten Falten ihrer l. Seite, die tief herabfallend sich der bereits gotischen Säule angleichen. Das Antik ist durchaus realistisch gehalten, doch in Porträtahnlichkeit durch die Umstände ausgeschlossen. Idealisiert dagegen erscheint das gefrönte Haupt (2 Säcken) das abgebrochen einer Reiterstatue am Dom zu Bamberg, in dem man Konrad III. vermutet (32). Der von einem schlanken Hals getragene, etwas rechts und aufwärts gewendete, von wallendem Haar umrahmte Kopf ist von außerordentlicher Kraft und Frische der Ausführung. — Unruhig wirkt gegenüber der Naumburger Statue das Doppelgräbn Heinrichs des Löwen und seiner Gemahlin Matilde im Braunschweiger Dom (31; f. 20). Die Figuren sind, obwohl sie liegen, doch wie stehend gehoben; andererseits aber fallen die Gewänder auch wieder nach dem Geize der Schwere; es erklärt sich die etwas ängstliche Faltengebung besonders des Frauengewandes. Man erkennt hieraus, daß der Künstler gewohnt und geübt war, Standbilder wie etwa 30 zu schaffen; hier sah er sich nun vor eine ganz neue Aufgabe gestellt, die er nicht widerprüflos zu lösen vermachte. Porträtahnlichkeit ist kaum anzunehmen. Heinrich trägt, wie üblich, in der Rechten das Modell seiner Stiftung, des Braunschweiger Domes.

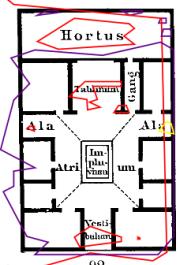
— 15 —



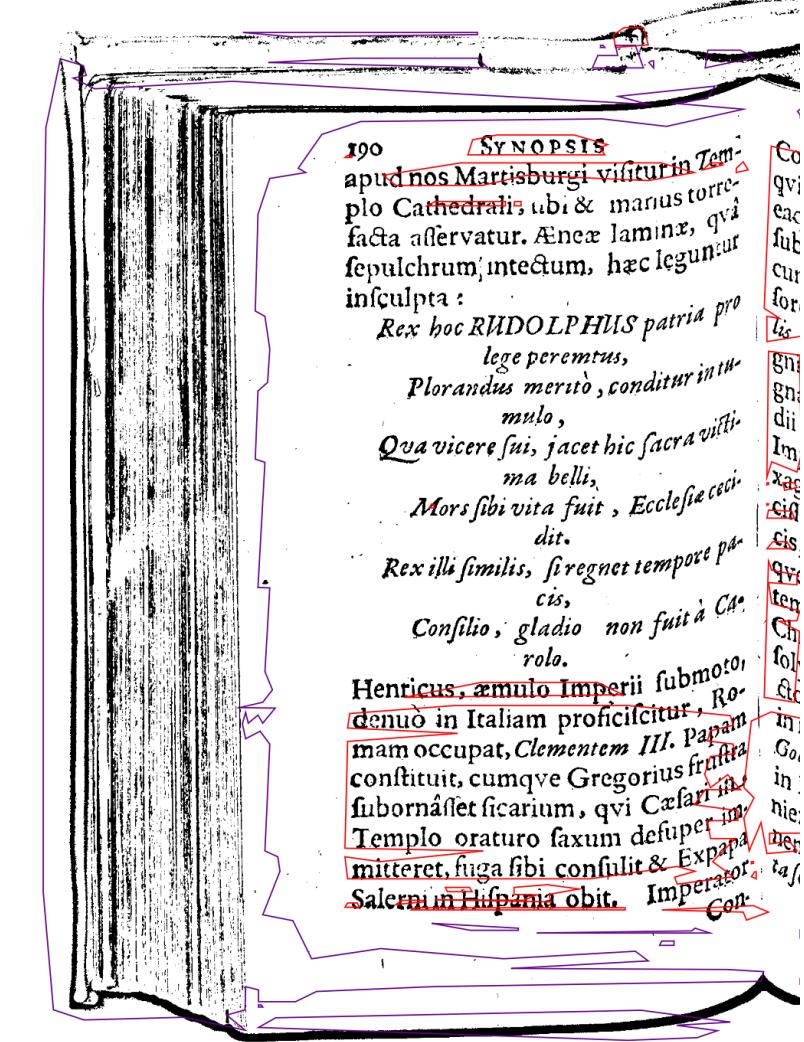
XVIII. Das römisch-griechische Haus.

Das altrömische Haus (92) war ursprünglich mit hohem Strohdach versehen; der Rauch des Herdes schwärzte den Mittelraum (daher Atrium) und fand seinen Ausweg durch die Tür und zwei seitliche Fügel, Aiae, welche Licht einließen. Das Aneinanderstoßen der Grundstücke in städtischen Beziehungen nötigte sodann, das Dach nach innen zu senken und den Regen durch eine von vier horizontalen Balken getragene Öffnung (Compluvium) in der Mitte des Atriums in ein Bassin (Impluvium) abzuleiten; die für die Lichtzufuhr leicht entbehrlichen Alae blieben. Dem Eingang (Vestibulum, Fauces) gegenüber befand sich, nach dem Atrium wie nach dem am Ende des Grundstückes liegenden Garten in voller Breite geöffnet, das Geschäftszimmer des Hausherrn, das Tablinum (tabulae). Ein niedriger Oberstock war Regel, zumal für das durch die Alae gebildete Querschiff. Bei größeren Abmessungen wurde das Compluvium durch vier Säulen geschnürt; aus dem Atrium corinthiscum wurde dann ein Atrium corinthicum.

Die griechische Hausanlage hatte sich unterdessen aus einem von dem Megaron beherrschten Hof, welchen offene Hallen und gedeckte Galässe umgaben, allmählich zu einem regelmäßigen Grundriss entwickelt: dem römischen Atrium entsprach das Peristyl, auf welches alle Räume des Hauses mündeten; das Megaron behielt seine Lage auf der dem Eingang genau überliegenden Seite, wurde aber auf die Bezugszimmer (Exedra), dem Tablinum entsprechend, in ganzer Breite nach dem Peristyl geöffnet und mit den übrigen Räumen unter einem Dache vereinigt. Bei größeren Häusern reihten sich zwei solcher peristylen Anlagen, vorne das Männerhaus, hinten das Frauenhaus, aneinander.



Druck von Karl Marquart in Leipzig.



190 **SYNOPSIS**
apud nos Martisburgi visitur in Templo Cathedrali ubi & manus torrefacta asservatur. Aeneæ laminæ, qvi sepulchrum intectum, hæc leguntur insculpta:
Rex hoc RUDOLPHUS patria pro lege peremtus,
Plorandus merito conditum in thymulo,
Qua vicere sui, jacet hic sacra vittima belli,
Mors sibi vita fuit, Ecclesie cecidit.
Rex illi similis, si regnet tempore pacis,
Confilio, gladio non fuit à Cæro.
Salerni in Hispania obit. Imperator

Henricus, amulo Imperii submotor denuo in Italiam proficiscitur Romanum occupat, Clementem III. Papam constitut, cumqve Gregorius frater subornasset sicarium, qvi Cæsar in Templo oraturo saxum desuper imitteret, fuga sibi consulit & Expatria Salerni in Hispania obit. Imperator

Fazit und Ausblick

- Lernt gut, generalisiert schlecht
- Ansatz (für dieses Projekt) sehr aussichtsreich
 - Trainingsdaten lassen sich sehr schnell erstellen
 - Material weitestgehend unproblematisch
- Weiter breit Trainingsdaten erstellen
- Gezieltes Active Learning basierend auf Qualitätssicherung (kommt gleich)
- Ggf. Regionentypen schrittweise wieder weiter ausdifferenzieren



Agenda

1. Allgemeine Projektinfos
2. Genereller (OCR) Workflow
3. Layoutanalyse
- 4. OCR-Qualitätssicherung und –Optimierung**
5. NER und NEL



Überblick

- Eingehende manuelle Prüfung aller Werke/Seiten nicht gangbar
- Stattdessen gezielte Prüfung, basierend auf Qualitätsabschätzung
 - Auf Werk-/Seitenebene → jeweils sortiert nach (vermuteter) „Qualität“
 - Flexibel und experimentell!
- Art des Eingriffs dann je nachdem
 - Anpassung von Parametern, Anwendung eines anderen Modells, ...
 - Gezieltes Nachtraining auf Werk-/Clusterebene (Active Learning!)
 - ...?
- Frage: Wie Qualität abschätzen?
 - Bereits gute Erfahrungen mit OCR-Konfidenz, allerdings auch klare Schwächen
 - Nur pragmatische (und idealerweise direkt einsetzbare) Lösungen sinnvoll



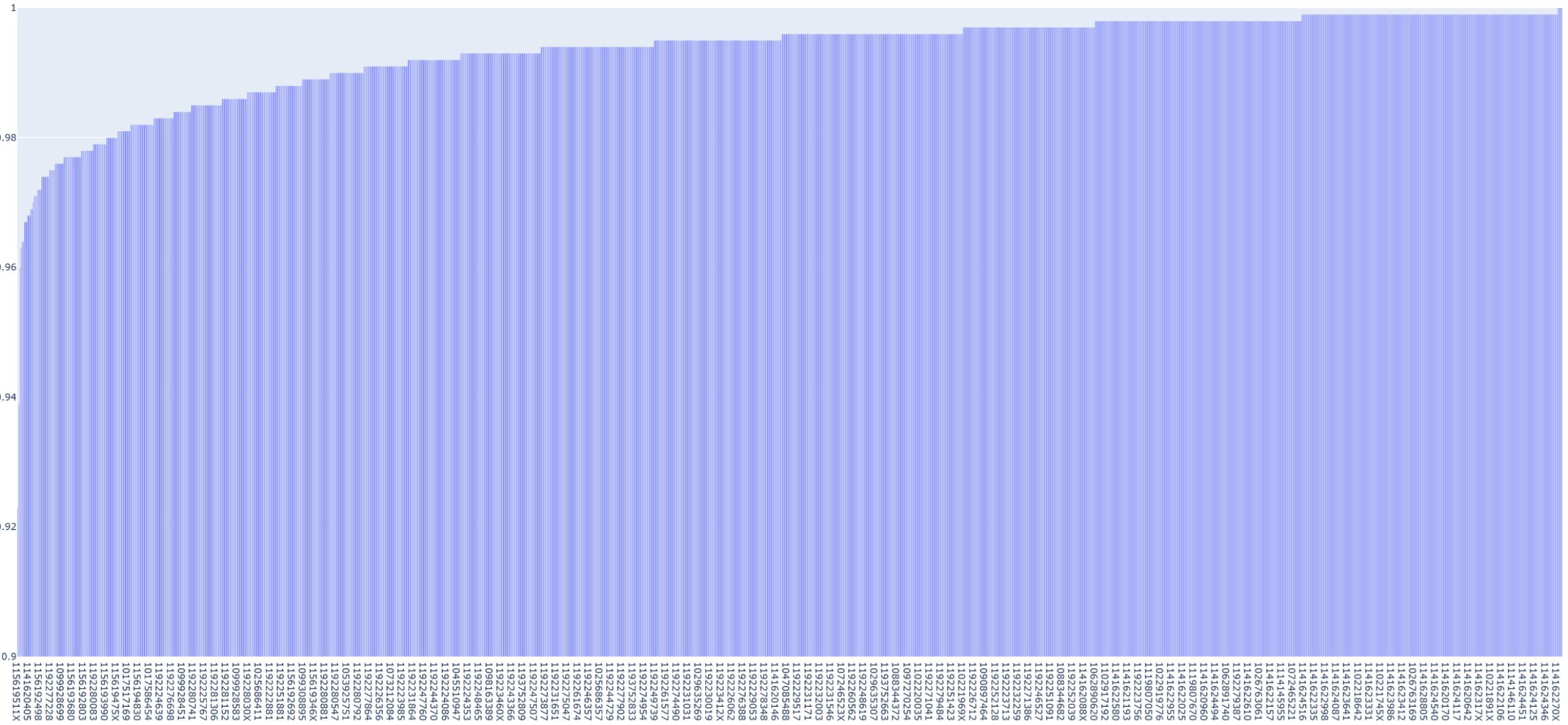
Rückblick: Anwendungsfall „Fotidis‘ Schundhefte“

- Material: Deutsche Heftromane, hauptsächlich 19. Jh.
- Ziel-Anwendung: Quantitative Analyse / Computational Literary Studies
→ (sehr) gute aber nicht perfekte Qualität erforderlich
- Workflow damals
 - Preprocessing und Layoutanalyse mit OCRopus (teils recht fehlerhaft)
 - Texterkennung mit Calamari
 - Erfassung der Calamari-Konfidenz; Aggregation auf Werk- und Seitenebene
 - Prüfung/Korrektur in LAREX
 - Seiten aufsteigend nach Konfidenz sortiert
 - Abbruch, sobald 2-3 Seiten (quasi) korrekt / gut genug
- Hocheffizient!
- Aber: Gerade Layoutanalyse mittlerweile deutlich weiter



Rückblick: Anwendungsfall „Fotidis‘ Schundhefte“

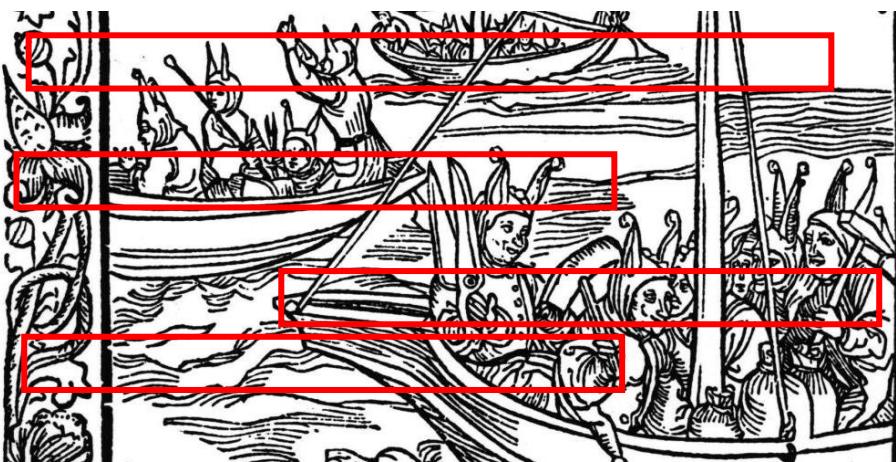
Average recognition confidence per book



Stärken/Schwächen OCR-Konfidenz

~~Text Text Text Text Text~~

teilsegmentierte Zeile, Zeilen in Bild
→ OCR-Konfidenz klarer Indikator



~~gubergren, no sea takimata sanctus est~~
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd

Übersehene Spaltentrennung
→ OCR-Konfidenz ggf. top,
Ergebnis jedoch Quatsch

Fazit und Ausblick

- Gezielte Korrektur aussichtsreich ...
- ... gerade im Zusammenspiel mit Active Learning der Layoutanalyse
- OCR-Konfidenz kann guter Indikator sein ...
- ... allerdings stark abhängig von Art der Fehler

- Einsatz von / Kombination mit NLP-Methoden naheliegend und vielversprechend

Agenda

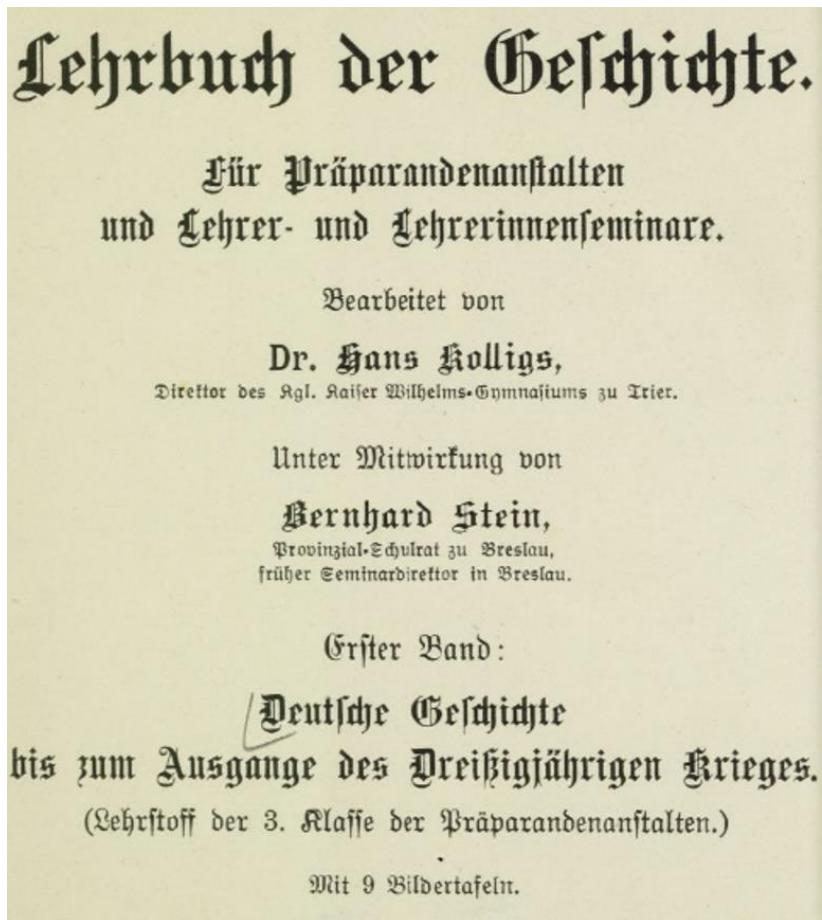
1. Allgemeine Projektinfos
2. Genereller (OCR) Workflow
3. Layoutanalyse
4. OCR-Qualitätssicherung und –Optimierung
- 5. NER und NEL**



Überblick

- Systematische Erfassung von Akteuren aus der historischen Schulbuchproduktion unter Einsatz von Named Entity Recognition (NER) und Named Entity Linking (NEL)
- Named Entity Recognition (NER)
 - Benannte Entitäten in natürlichsprachigen Texten automatisiert erkennen und in vordefinierte Kategorien einordnen
 - Standard-Unterscheidung in vier Klassen: Person (PER), Location (LOC), Organisation (ORG), Miscellaneous (MISC); je nach Modell auch andere Klassifizierung möglich
- Named Entity Linking (NEL)
 - Verlinkung/Verknüpfung von Entitäten zu anderen Quellen, z. B. einer Normdatei oder einer anderen zuvor festgelegten Wissensbasis (GND, Wikipedia, ...)
 - Entität eindeutig identifizieren und gegenüber gleichlautenden Entitäten disambiguieren

Beispiel NER-Ergebnis auf fehlerhafter OCR



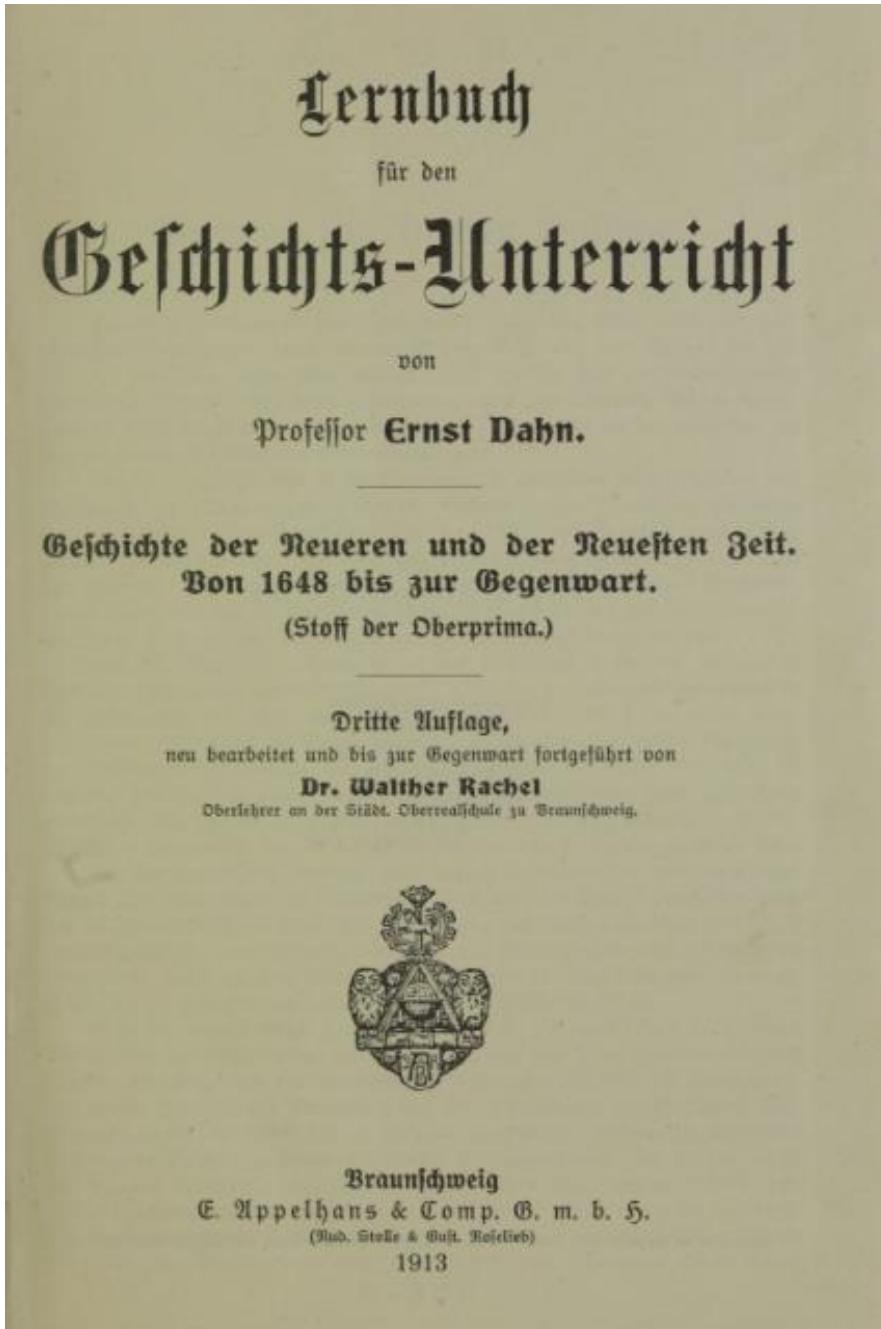
Lehrbuch der Geschichte MISC .
Für Präparandenanstalten
und Lehrer- und Lehrerinnenseminare.
Bearbeitet von
Dr. Hans Kolligs PER .
Direktor des Kal ORG . Kaiser Wilhelms-Gymnasiums ORG zu Ttier LOC .
Unter Mitwirkung von
Bernhard Stein PER .
Provinziol. Schulrat zu Breslau LOC ,
früher Seminardirektor in Greslas LOC .
Erster Band:
Deutsche Geschichte
bis zum Ausgange des Dreißigjährigen Krieges MISC .
(Lehrstoff der 3. Klasse der Präparandenanstalten.)
Mit 9 Bildertafeln.



Workflow

- Plain Text erfassen
 - Mittels OCR erzeugen
 - Aus bestehenden Ergebnissen extrahieren (für Teilkorpus liegen alte ABBYY Ergebnisse vor)
- Fokus auf Textabschnitte, die besonders viele relevante Entitäten enthalten
 - Titelblätter, Vorworte, ...
 - Gezielte Durchführung, basierend auf Infos aus METS
- NER: Modelle identifizieren Entitäten
- NEL: Unterstützte Disambiguierung und Verknüpfung mit externen Quellen (GND)
- Verknüpfung mit den Ursprungsdaten (ALTO)
- (Visualisierung, ggf. Netzwerkanalyse etc.)





00000005.xml

| PcGts | Page | ReadingOrder |
|-------|------|---|
| 1 | | <?xml version='1.0' encoding='utf-8'?> |
| 2 | | <PcGts xmlns="http://schema.primaresearch.org/PAGE/gts/pagecontent/2019-07-15" |
| 2 | | xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" |
| 2 | | xsi:schemaLocation="http://schema.primaresearch.org/PAGE/gts/pagecontent/2019-07-15 |
| 2 | | http://schema.primaresearch.org/PAGE/gts/pagecontent/2019-07-15/pagecontent.xsd"> |
| 3 | | <Metadata> |
| 4 | | <Creator>xml_converter</Creator> |
| 5 | | <Created>2024-10-01T10:09:55.392561</Created> |
| 6 | | <LastChange>2024-10-01T10:09:55.395534</LastChange> |
| 7 | | </Metadata> |
| 8 | | <Page imageFilename="00000005.png" imageWidth="2254" imageHeight="3618"> |
| 9 | | <ReadingOrder> |
| 10 | | <OrderedGroup id="g0"> |
| 11 | | <RegionRefIndexed index="0" regionRef="r0"/> |
| 12 | | <RegionRefIndexed index="1" regionRef="r1"/> |
| 13 | | <RegionRefIndexed index="2" regionRef="r2"/> |
| 14 | | <RegionRefIndexed index="3" regionRef="r3"/> |
| 15 | | <RegionRefIndexed index="4" regionRef="r4"/> |
| 16 | | <RegionRefIndexed index="5" regionRef="r5"/> |
| 17 | | <RegionRefIndexed index="6" regionRef="r6"/> |
| 18 | | <RegionRefIndexed index="7" regionRef="r7"/> |
| 19 | | </OrderedGroup> |
| 20 | | </ReadingOrder> |
| 21 | | <TextRegion id="r0" type="paragraph"> |
| 22 | | <Coords points="696,338 1246,348 1246,508 1074,562 1074,619 1858,692 1858,908 |
| 22 | | 1026,1012 1026,1043 1372,1155 1362,1237 586,1227 586,1155 922,1043 922,1012 94,908 |
| 22 | | 94,692 872,619 872,562 686,508 686,348"/> |
| 23 | | <TextLine id="r0_10"> |
| 24 | | <Coords points="696,352 772,352 784,388 836,388 844,384 904,384 912,384 980,384 |
| 24 | | 996,352 1048,352 1060,384 1132,384 1144,348 1236,348 1144,348 1236,348 1236,508 |
| 24 | | 1144,508 1236,508 1144,508 1132,472 1060,472 1048,476 996,476 980,476 912,476 904,476 |
| 24 | | 844,476 836,476 784,476 772,508 696,508 696,352"/> |
| 25 | | <Baseline points="696,472 1236,472"/> |
| 26 | | <TextEquiv> |
| 27 | | <Unicode>Kernbuch</Unicode> |
| 28 | | </TextEquiv> |
| 29 | | </TextLine> |
| 30 | | <TextLine id="r0_11"> |
| 31 | | <Coords points="882,562 900,562 902,564 930,564 933,576 953,576 954,562 985,562 |
| 31 | | 986,562 1010,562 1016,575 1034,575 1038,575 1064,575 1064,606 1038,606 1034,607 |
| 31 | | 1016,607 1010,607 986,607 985,607 954,607 953,607 933,607 930,607 902,607 900,619 |
| 31 | | 882,619 882,562"/> |
| 32 | | <Baseline points="882,607 1064,607"/> |
| 33 | | <TextEquiv> |
| 34 | | <Unicode>für den</Unicode> |
| 35 | | </TextLine> |

Kernbuch
für den
Oeschichts-Unterricht
von
Professor Ernst Dahn.
Geschichte der Neueren und der Neuesten Zeit. Von 1648 bis zur Gegenwart.
(Stoff der Oberprima.)
Dritte Auflage,
neu bearbeitet und bis zur Gegenwart fortgeführt von Dr. ttialtber Kachel
Oberlehrer an der Stöbt. Oberrealschule zu Braunschroeig.
Braunschweig
E. Appelhans & Comp. 05. m. b. H.
(Rud. Stolle & Gust. Roselieb)
1913

Kernbuch
für den
Oeschichts-Unterricht
von
Professor Ernst Dahn PER gnd .
Geschichte der Neueren und der Neuesten Zeit. Von 1648 bis zur Gegenwart.
(Stoff der Oberprima.)
Dritte Auflage,
neu bearbeitet und bis zur Gegenwart fortgeführt von Dr PER , ttialtber Kachel PER
Oberlehrer an der Stöbt. Oberrealschule ORG zu Braunschroeig LOC gnd .
Braunschweig E. Appelhans & Comp ORG . 05. m. b. H.
(Rud PER . Stolle PER & Gust PER . Roselieb PER)
1913



Erste kleine Experimente

- Aktuell basierend auf bereits vorliegenden ABBYY Ergebnissen, mit eher überschaubarer Qualität
- Testen verschiedener Standard-Modelle
 - Spacy https://spacy.io/models/de#de_core_news_lg
 - Flair <https://github.com/flairNLP/flair>
 - Babelscape <https://huggingface.co/Babelscape/wikineural-multilingual-ner>
- Erste kleine Auswertung
 - Menschlich erzeugte Golddaten mit Fokus auf Personen und Orte (173 Entitäten aus drei Werken)
 - Datengrundlage sind die Scans, nicht das erzeugte OCR
 - Flair Modell liefert beste Ergebnisse (Precision >0,9; Recall >0,8)
 - Exakte Evaluation schwierig → grobe Werte als erste Abschätzung!

Fazit und Ausblick

- Noch einige Herausforderungen, u. a.
 - Teils sehr fehlerhafte OCR
 - Umgang mit langen zusammengesetzten Entitäten
„Frau Akademische Oberrätin Priv. Doz. Dr. habil. phil. Katrin Dennerlein (geb. Fischer)“
 - Disambiguierung (Herr Müller ≠ Herr Müller ≠ Herr Müller)
- Prinzipiell aber erstaunlich robust, performant und aussichtsreich
- Verschiedene Möglichkeiten, Ergebnisse erheblich zu verbessern, u. a.
 - Anwendung auf bessere OCR
 - Bessere NER Modelle, ggf. durch Finetuning / Domain Adaptation
 - Ergänzung durch heuristische Methoden
 - NEL erheblich unterstützten (GND-API + eigene GUI)



Vielen Dank für die Aufmerksamkeit!

Projektbeteiligte:

- Fabian Dombrowski (NER und NEL, GEI)
- Janik Haitz (Layoutanalyse und Texterkennung, ZPD)
- Corinna Keupp (NER und NEL, ZPD)
- Florian Langhanki (Layoutanalyse und Texterkennung, ZPD)
- Johannes Leitgeb (NER und NEL, ZPD)
- Sebastian Klaes (Koordination, GEI)
- Anke Hertling (Antragstellerin, GEI)
- Christian Reul (Antragsteller, ZPD)

