Vierter DTA-Workshop

Aufbau historischer Sprachressourcen: Arbeiten mit den Angeboten des Deutschen Textarchivs

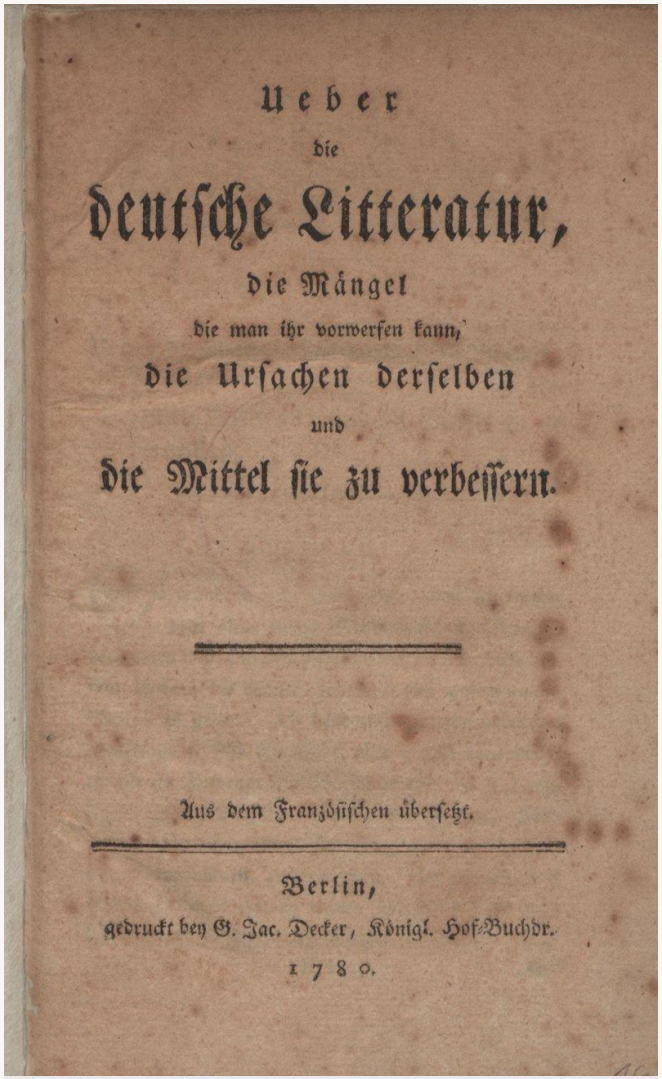
Gliederung

- Textressourcen im WWW:
  Probleme aus der Nutzerperspektive
- Das Deutsche Textarchiv (DTA):
  Kurzvorstellung
- Vierter DTA-Workshop:
  Einführung in das Programm

Problem 1: Verstreute Ressourcen

"Wir beſitzen noch keine von der ganzen Nation gebilligte Sammlung, in der man alle Worte und Redensarten fände, nach denen man die Reinigkeit der Sprache ſicher beurtheilen könnte."

Friedrich II.: Über die deutsche Literatur [...].
Aus dem Französischen übersetzt.
Berlin, 1780, S. 6. Online-Version:
http://www.deutschestextarchiv.de/friedrich_literatur_1780/40

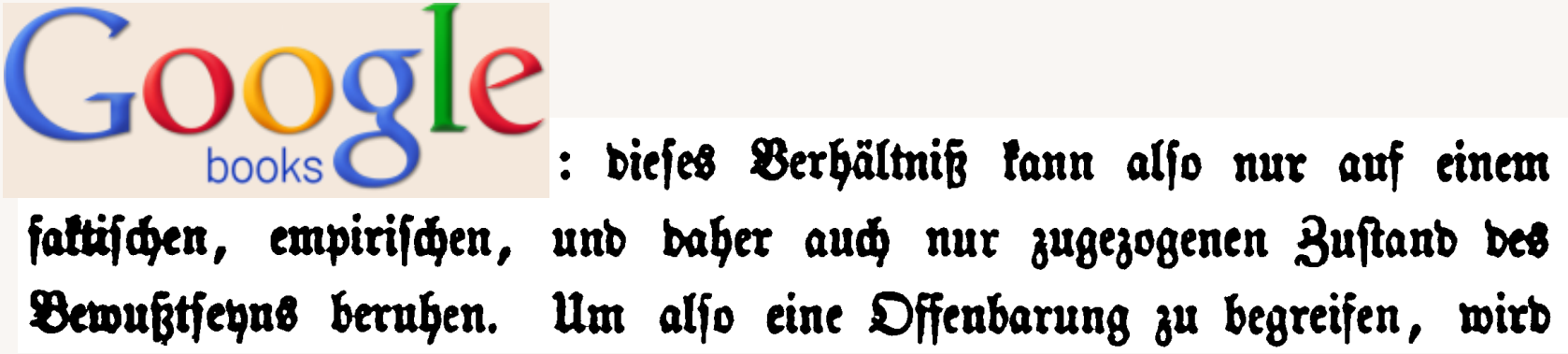

Ueber
die
deutſche Litteratur,
die Mängel
die man ihr vorwerfen kann,
die Urſachen derſelben
und
die Mittel ſie zu verbeſſern.
Aus dem Franzöſiſchen überſetzt.
Berlin,
gedruckt bey G. Jac. Decker, Königl. Hof-Buchdr.
1780.

Nachweis: Berlin SBB-PK, Bibl. Diez oct. 8354

Problem 1: Verstreute Ressourcen

Problem 1: Verstreute Ressourcen

Problem 2: Fehlerhafte Texte

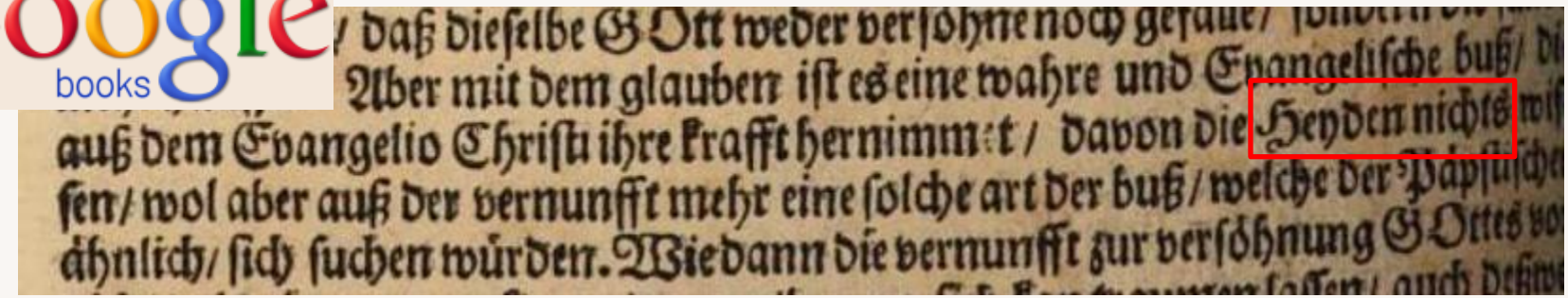


[…]: dieses Verhältniß kann also nnr ans einem saktischen, empirischen, nnd daher anch nnr zngezogenen Znstand des Bewnßtseyns bernhen. […]

Schellings sämmtliche Werke. Hrsg. v. K. F. A. Schelling. 2. Abtheilung, 3. Band. Stuttgart u.a. : Cotta 1858, S. 185. http://books.google.de/books?id=32-sknQV_iYC&hl=de&pg=PA185

- fehlerhafte Zeichen: 11/130 (≈ 91,5 % accuracy)
- fehlerhafte Tokens: 10/18 (≈ 44,4% accuracy)

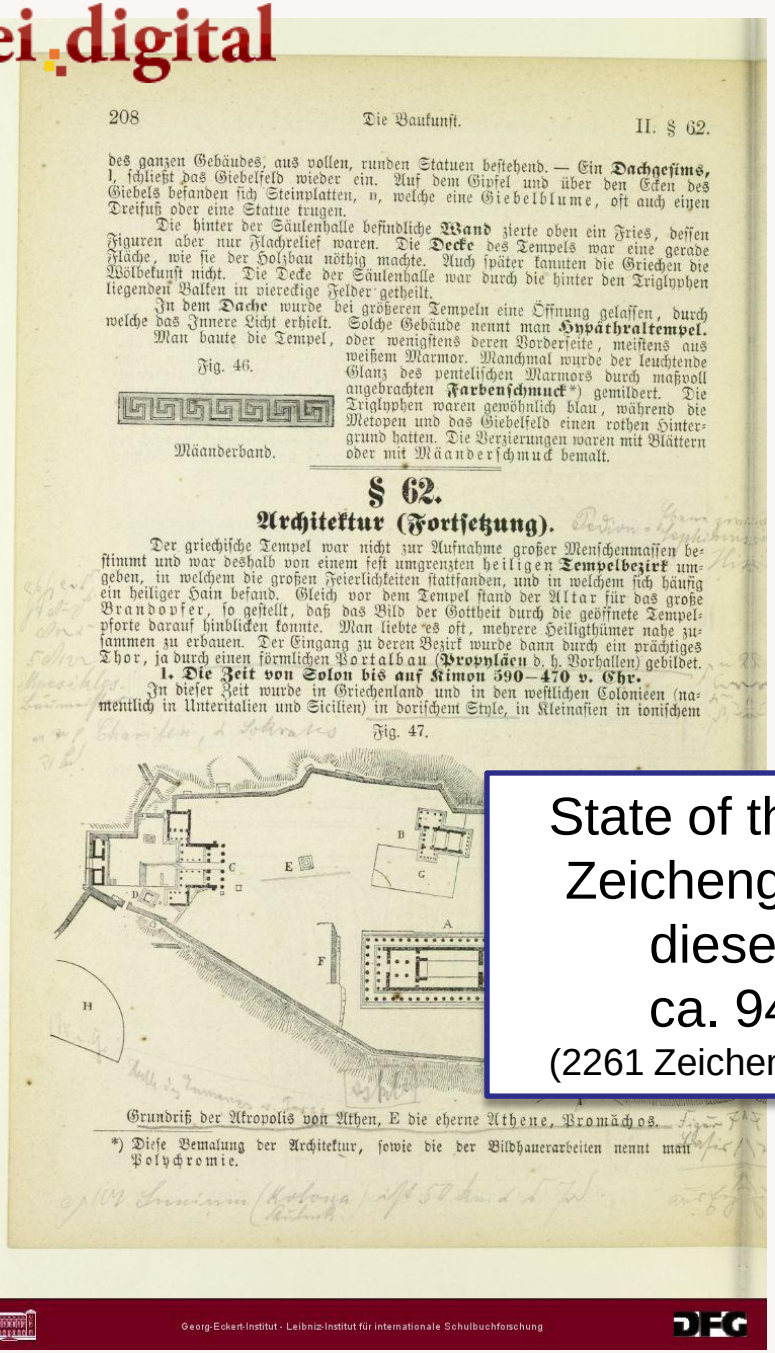
Problem 2: Fehlerhafte Texte



[…] Aber mit dem glauben ifkeseinewahre und Evangelifehe bufil dll WE dm! Evqngel-o Chrifli ihre krafft hernimmxtr davon die-**Hydraulik**!? will fini- ijool aber auli der vernunffc mehr eine folcloearcder buß/welcdedkkPäyfillÜil ähulkcb/ [ich "nchen würdem. […]

Philipp Jacob Spener: Die evangelische Glaubensgerechtigkeit, von Johann Brevings Angriffen gerettet. Frankfurt a. M.: Zunner, **1684**. http://books.google.de/books?id=HF9KAAAAcAAJ&pg=PA826

(NB: 1st Hit for "**Hydraulik**" in Google Books Corpus)



208 Die Baukunst. II. § 62.

des ganzen Gebäudes, aus vollen, runden Statuen bestehend. — Ein **Dachgesims**, 1, schließt das Giebelfeld wieder ein. Auf dem Gipfel und über den Ecken des Giebels befanden sich Steinplatten, n, welche eine Giebelblume, oft auch einen Dreifuß oder eine Statue trugen.

Die hinter der Säulenhalle befindliche **Wand** zierte oben ein Fries, dessen Figuren aber nur Flachrelief waren. Die **Decke** des Tempels war eine gerade Fläche, wie sie der Holzbau nöthig machte. Auch später kannten die Griechen die Wölbekunst nicht. Die Decke der Säulenhalle war durch die hinter den Triglyphen liegenden Balken in viereckige Felder getheilt.

In dem **Dache** wurde bei größeren Tempeln eine Öffnung gelassen, durch welche das Innere Licht erhielt. Solche Gebäude nennt man **Hypäthraltempel.** Man baute die Tempel, oder wenigstens deren Vorderseite, meistens aus weißem Marmor. Manchmal wurde der leuchtende Glanz des pentelischen Marmors durch maßvoll angebrachten **Farbenschmuck***) gemildert. Die Triglyphen waren gewöhnlich blau, während die Metopen und das Giebelfeld einen rothen Hintergrund hatten. Die Verzierungen waren mit Blättern oder mit Mäanderschmuck bemalt.

Fig. 46.

Mäanderband.

§ 62.

Architektur (Fortsetzung).

Der griechische Tempel war nicht zur Aufnahme großer Menschenmassen bestimmt und war deshalb von einem fest umgrenzten heiligen **Tempelbezirk** umgeben, in welchem die großen Feierlichkeiten stattfanden, und in welchem sich häufig ein heiliger Hain befand. Gleich vor dem Tempel stand der Altar für das große Brandopfer, so gestellt, daß das Bild der Gottheit durch die geöffnete Tempelpforte darauf hinblicken konnte. Man liebte es oft, mehrere Heiligthümer nahe zusammen zu erbauen. Der Eingang zu deren Bezirk wurde dann durch ein prächtiges Thor, ja durch einen förmlichen Vorbau (**Propyläen** d. h. Vorhallen) gebildet.

1. Die Zeit von Solon bis auf Kimon 590—470 v. Chr.

In dieser Zeit wurde in Griechenland und in den westlichen Colonieen (namentlich in Unteritalien und Sicilien) in dorischem Style, in Kleinasien in ionischem

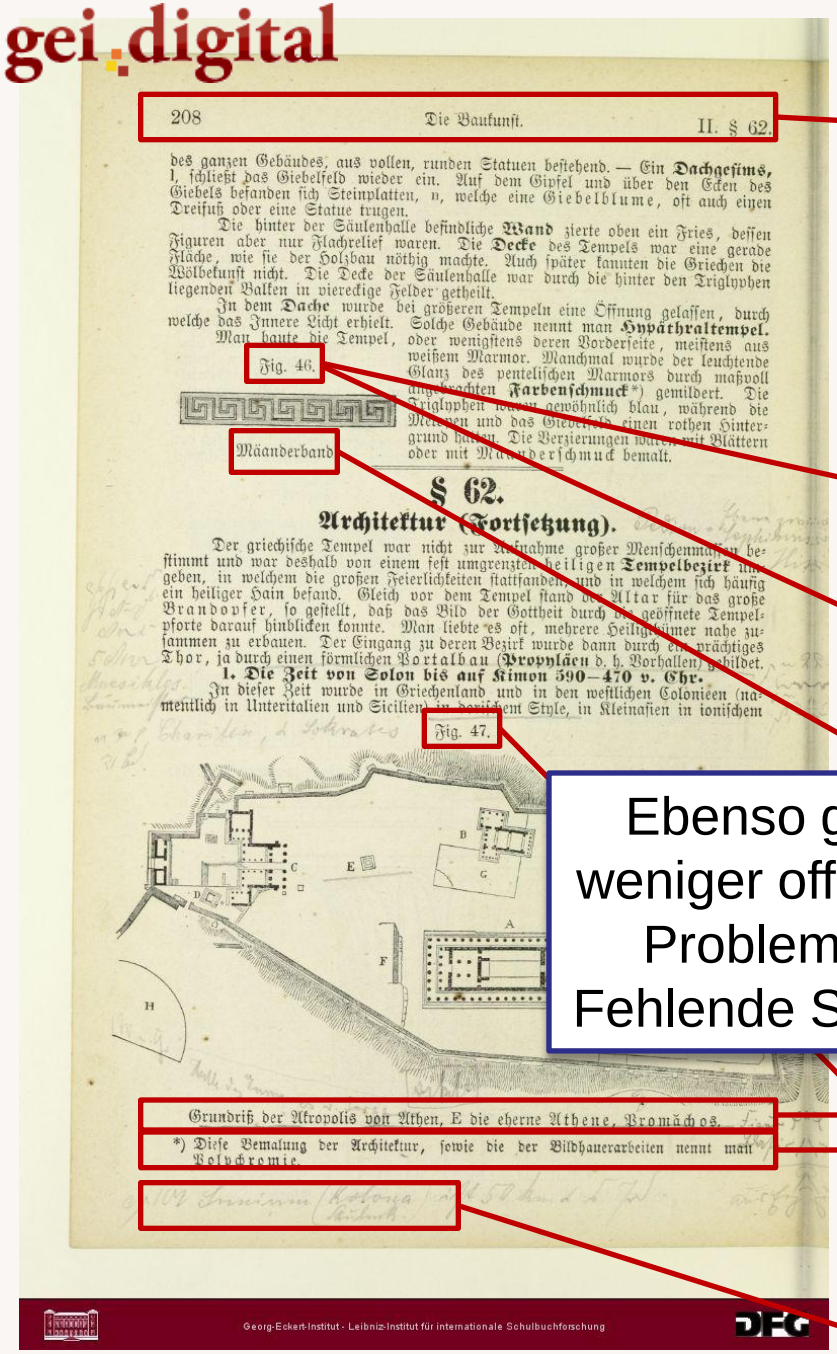
Fig. 47.

Grundriß der Akropolis von Athen, E die eherne Athene, Promächos.

*) Diese Bemalung der Architektur, sowie die der Bildhauerarbeiten nennt man Polychromie.

State of the art OCR;
Zeichengenauigkeit
dieser Seite:
ca. 94,958%
(2261 Zeichen, 114 fehlerhaft)

Grundriß der Akropolis von Athen, E die eherne Athene, Promnrih n3.
*) Diese Bemalung der Architektur, sowie die der Bildhauerarbeiten nennt rnttft : - n Polychromie.
208 Die Baukunst. II § 62
Mirnietl ^^udes aus vollen, runden Statuen bestehend. - Ein Dachaesims. as. ©tebelfclb roxcber ein. Auf dem Gipfel und über den Ecken des Giebels befanben st* Steinplatten, n, welche eine Giebelblume, oft auch einen Dreifuß oder eine Statue trugen. ' 1 ;
Die hinter der Säulenhalle befinbliche Wand zierte oben ein Fries beffen Figuren aber nur Flachrelief waren. Die Decke bes Tempels war eine' aerobe ^te [*£ beL.^°ä ^thtg machte. Auch später kannten bte Griechen die Wolbekunst mcht. Die Decke der ^anlenhalle war durch die hinter den Triglyphen liegenden Balken m viereckige Felber getheilt.
_ dem Dache wurde bei größeren Tempeln eine Öffnung gelassen, durch welche das -jnnere Licht erhielt. Solche Gebäude nennt man Hypathraltempel. Man baute bte Tempel, ober wenigstens beren Vorderseite, meistens aus 4ß weißem Marmor. Manchmal würbe ber leuchteube
tir9' ' ^lanZ bes pentelischen Marmors burch maßvoll angebrachten Farbenschmuck*) gemilbert. Die Triglyphen waren gewöhnlich blau, währenb bie Metopen unb bas Giebelfeld einen rothen Hinter- (m.. ^ r c grunb hatten. Die Verzierungen waren mit Blättern
nit Mäanderschmuck bemalt.
ung).
pel war nicht zur Aufnahme großer Menschenmassen
shalb von einem fest umgrenzten heiligen
en, in welchem bie großen Feierlichkeiten stattfanben,
häufia ein heiliger Ham befanb. Gleich vor dem Tempel
s große Branbopfer, fo gestellt, baß bas Bilb ber
öffnete Tempelpforte baranf Hinblicken konnte. Man liebte es oft, mehrere Heiligthümer nahe zusammen zu erbauen. Der Eingang zu bereit Bezirk würbe bann burch ein prächtiges Thor, ja burch einen förmlichen Portalb au (Propyläen b. h. Vorhallen) aebilbet.
1* Dte Zeit von Solon bis auf Kimon 390—470 v. Chr.
In biefer Zeit würbe in Griechenland) unb in bett westlichen Colonieen (namentlich in Unterttalien unb Sicilien) in borischem Style, in Kleinasien in ionischem Fig' 47.
x S'%%iM!SSlÄi//iiiiiiiiii////Z',,



208 Die Baukunst. II. § 62.

des ganzen Gebäudes, aus vollen, runden Statuen bestehend. — Ein **Dachgesims**, l, schließt das Giebelfeld wieder ein. Auf dem Gipfel und über den Ecken des Giebels befanden sich Steinplatten, n, welche eine Giebelblume, oft auch einen Dreifuß oder eine Statue trugen.

Die hinter der Säulenhalle befindliche **Wand** zierte oben ein Fries, dessen Figuren aber nur Flachrelief waren. Die **Decke** des Tempels war eine gerade Fläche, wie sie der Holzbau nöthig machte. Auch später kannten die Griechen die Wölbekunst nicht. Die Decke der Säulenhalle war durch die hinter den Triglyphen liegenden Balken in viereckige Felder getheilt.

In dem **Dache** wurde bei größeren Tempeln eine Öffnung gelassen, durch welche das Innere Licht erhielt. Solche Gebäude nennt man **Hypäthraltempel.** Man baute die Tempel, oder wenigstens deren Vorderseite, meistens aus weißem Marmor. Manchmal wurde der leuchtende Glanz des pentelischen Marmors durch maßvoll angebrachten **Farbenschmuck***) gemildert. Die Triglyphen waren gewöhnlich blau, während die Metopen und das Giebelfeld einen rothen Hintergrund hatten. Die Verzierungen waren mit Blättern oder mit Mäanderschmuck bemalt.

Fig. 46.

Mäanderband

§ 62.

Architektur (Fortsetzung).

Der griechische Tempel war nicht zur Aufnahme großer Menschenmassen bestimmt und war deshalb von einem fest umgrenzten heiligen **Tempelbezirk** umgeben, in welchem die großen Feierlichkeiten stattfanden, und in welchem sich häufig ein heiliger Hain befand. Gleich vor dem Tempel stand der **Altar** für das große Brandopfer, so gestellt, daß das Bild der Gottheit durch die geöffnete Tempelpforte darauf hinblicken konnte. Man liebte es oft, mehrere Heiligthümer nahe zusammen zu erbauen. Der Eingang zu deren Bezirk wurde dann durch ein prächtiges Thor, ja durch einen förmlichen **Vortalbau** (**Propyläen** d. h. Vorhallen) gebildet.

1. Die Zeit von Solon bis auf Kimon 590—470 v. Chr.

In dieser Zeit wurde in Griechenland und in den westlichen Colonieen (namentlich in Unteritalien und Sicilien) in dorischem Style, in Kleinasien in ionischem

Fig. 47.

Grundriß der Akropolis von Athen, E die eherne Athene, Promächos.

*) Diese Bemalung der Architektur, sowie die der Bildhauerarbeiten nennt man Polychromie.

Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung

DFG

Ebenso großes, nur weniger offensichtliches Problem bei OCR: Fehlende Strukturierung

Grundriß der Akropolis von Athen, E die eherne Athene, Promnrih n3.

*) Diese Bemalung der Architektur, sowie die der Bildhauerarbeiten nennt rnttft : - n Polychromie.

208 Die Baukunst. II § 62

Mirnietl ^^udes aus vollen, runden Statuen bestehend. - Ein Dachaesims, as. ©tebelfclb roxcber ein. Auf dem Gipfel und über den Ecken des Giebels befanben st* Steinplatten, n, welche eine Giebelblume, oft auch einen Dreifuß oder eine Statue trugen. ' 1 ;

Die hinter der Säulenhalle befinbliche Wand zierte oben ein Fries beffen Figuren aber nur Flachrelief waren. Die Decke bes Tempels war eine' aerobe ^te [*£ beL.^°ä ^thtg machte. Auch später kannten bte Griechen die Wolbekunst mcht. Die Decke der ^anenhalle war durch die hinter den Triglyphen liegenden Balken m viereckige Felber getheilt.

_ dem Dache wurde bei größeren Tempeln eine Öffnung gelassen, durch welche das -jnnere Licht erhielt. Solche Gebäude nennt man Hypathraltempel. Man baute bte Tempel, ober wenigstens beren Vorderseite, meistens aus 4ß weißem Marmor. Manchmal würbe ber leuchteube

tir9' ' ^lanZ bes pentelischen Marmors burch maßvoll angebrachten Farbenschmuck*) gem lbert. Die Triglyphen waren gewöhnlich blau, währenb bie Metopen unb bas Giebelfeld einen rothen Hinter- (m.. ^ r c grunb hatten. Die Verzierungen waren mit Blättern

nit Mäanderschmuck bemalt.

ung).

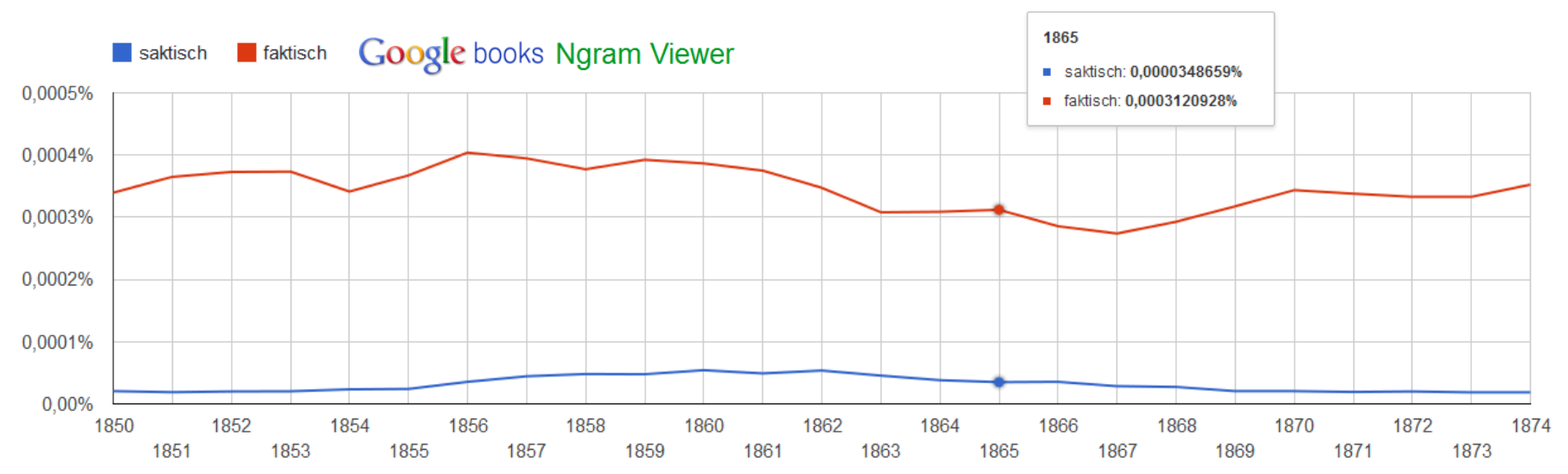
pel war nicht zur Aufnahme großer Menschenmassen shalb von einem fest umgrenzten heiligen en, in welchem bie großen Feierlichkeiten stattfanben, häufig ein heiliger Ham befanb. Gleich vor dem Tempel s große Branbopfer, fo gestellt, baß bas Bilb ber Gottheit burch bie geöffnete Tempel bforte baranf Hinblicken konnte. Man liebte es oft, mehrere Heiligthümer nahe zusammen zu erbauen. Der Eingang zu bereit Bezirk würbe bann burch ein prächtiges Thor, ja burch einen förmlichen Portalb au (Propyläen b. h. Vorhallen) aebilbet.

1* Die Zeit von Solon bis auf Kimon 390—470 v. Chr.

In biefer Zeit würbe in Griechenland) unb in bett westlichen Colonieen (namentlich in Unterttalien unb Sicilien) in borischem Style, in Kleinasien in ionischem Fig' 47.

x S'%%iM!SSlÄi//iiiiiiiiii////Z',,

Problem 2: Fehlerhafte Texte



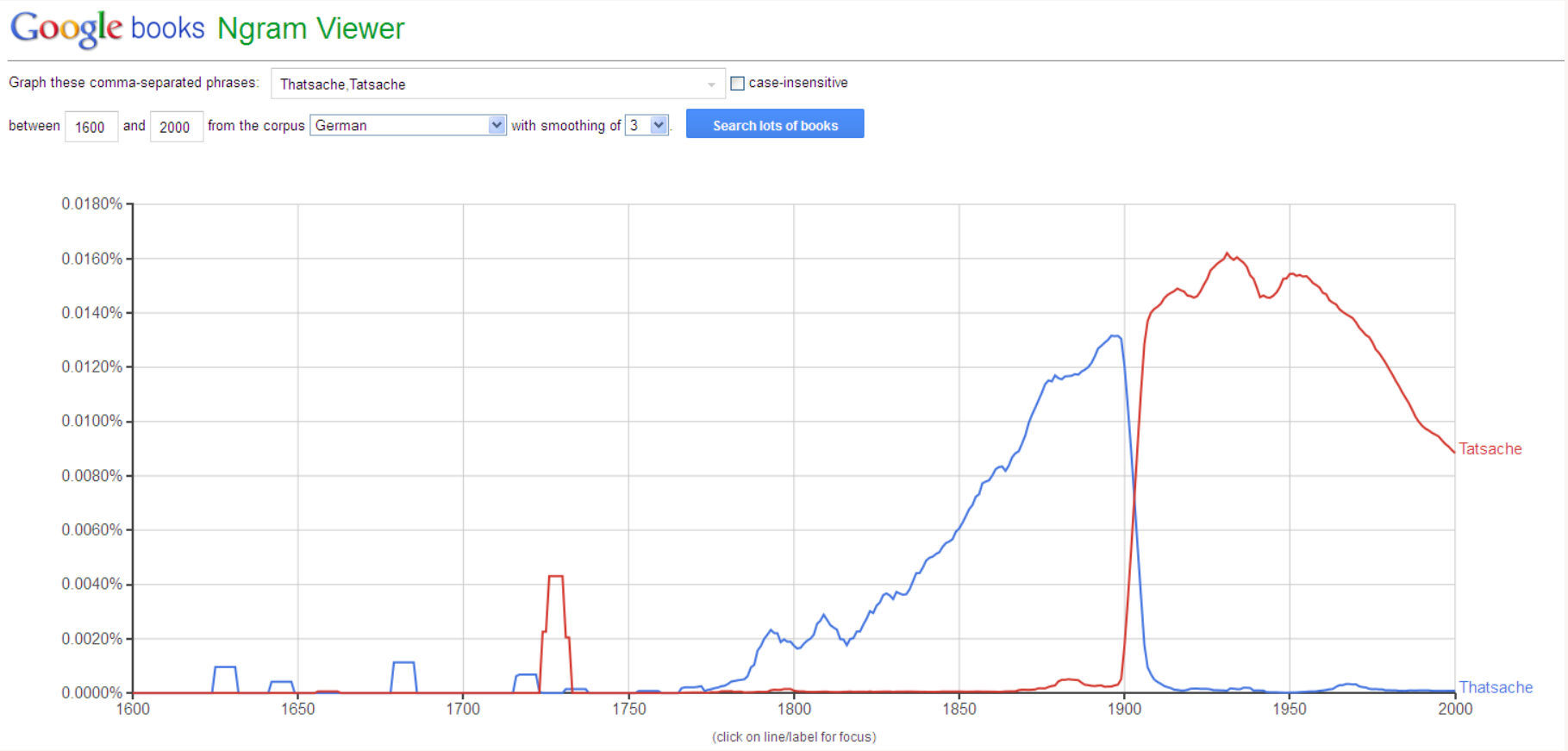
Pfeifer: EtymWB (via DWDS): *[…] faktisch Adj. ‘tatsächlich, wirklich’ (**Ende 18. Jh.**), meist adverbial gebraucht […]*

frühester Beleg "faktisch" in GB: 1729, http://bit.ly/15pSELY

frühester Beleg "factiſch" im DTA: 1790, http://bit.ly/1dGxKbJ

Problem 3: Historische Varianz

- fehlende Normierung der historischen Schreibvarianten



Problem 3: Historische Varianz

DTA Query Lizard

Query Term(s): "Befehl"

New Search | Query Lizard | Help

Base Query

Query Term(s): Befehl [Get Expansions]

Expander: Token

Expansions:

☑ BEfehlen	☑ Befelche	☑ Befhlen	☑ befele
☑ Befehl	☑ Befelchen	☑ Befihl	☑ befeles
☑ Befehlch	☑ Befelchs	☑ Befihle	☑ befelh
☑ Befehle	☑ Befele	☑ Befähl	☑ befellen
☑ Befehlen	☑ Befeleh	☑ Beuelch	☑ beuelch
☑ Befehles	☑ Befelen	☑ Bevehle	☑ bevehl
☑ Befehln	☑ Befelh	☑ befehl	☑ bevel
☑ Befehls	☑ Befels	☑ befehles	
☑ Befel	☑ Beferhl	☑ befehls	
☑ Befelch	☑ Beffehle	☑ befel	

Enable All | Disable All | Search

New Search | Query Lizard | Help

jurish@bbaw.de

DTA Query Lizard version 0.03

0.157955 sec

Problem 4: Zweifelhafte Vorlagen

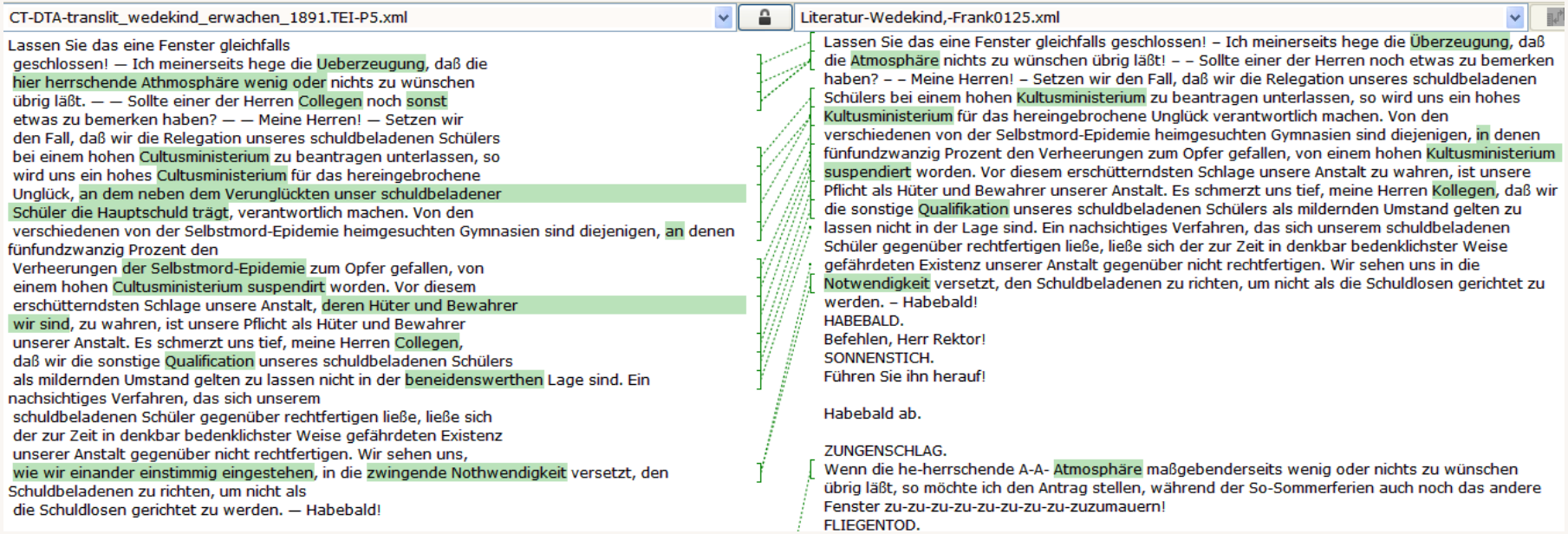
Zeno.org
MEINE BIBLIOTHEK

TextGrid
Repository

Problem 4: Zweifelhafte Vorlagen

Kollation zweier Transkriptionen (DTA::TextGridRep)

juxta
www.juxtasoftware.org



Wedekind, Frank: Frühlings Erwachen. Zürich, 1891. In: Deutsches Textarchiv http://www.deutschestextarchiv.de/wedekind_erwachen_1891, abgerufen am 17.09.2013.

Dass. in: Werke in drei Bänden. Berlin u. Weimar : Aufbau, 1969. In: TextGridRep, PID:hdl:11858/00-1734-0000-0005-94F9-5 abgerufen am 17.09.2013.

Problem 4: Zweifelhafte Vorlagen

Kollation zweier Transkriptionen (DTA::TextGridRep)

juxta
www.juxtasoftware.org

Brod] Brot

Thüre] Türe

Kniee] Knie

gieb] gib

Klex] Klecks

examiniren] examinieren

citiren] zitieren

Epheukranz] Efeukranz

Guitarre] Gitarre

Corset] Korsett

Kawboy] Cowboy

behülflich] behilflich

gräuliche] greuliche

Verabscheuenswürdiges] Verabscheuungswürdiges

unverrichteter Sache] unverrichtetersache

zu Muthe] zumute

Problem 4: Zweifelhafte Vorlagen

Kollation zweier Transkriptionen (DTA::TextGridRep)

juxta
www.juxtasoftware.org

Weiß der Himmel, mir ist um mich nicht mehr bange.
Ich habe diesmal zu tief hinunterblickt.]

Weiß der Himmel, mir ist um mich nicht mehr bange.
Dazu habe ich diesmal zu tief hinuntergeblickt.

Wir sehen uns, wie wir einander einstimmig eingestehen, in die zwingende Nothwendigkeit versetzt, den Schuldbeladenen zu richten.

so wird uns ein hohes Cultusministerium für das hereingebrochene Unglück, an dem neben dem Verunglückten unser schuldbeladener Schüler die Hauptschuld trägt, verantwortlich machen.

Gliederung

- Textressourcen im WWW : Probleme aus der Nutzerperspektive
- Das Deutsche Textarchiv (DTA): Kurzvorstellung
- Vierter DTA-Workshop – Einführung in das Programm

Deutsches Textarchiv (DTA)

Anmelden (DTAQ)

DTA

Texte ▼ Projekt ▼ Dokumentation ▼ Impressum

suchen

in den Titeldaten im Korpus auf der Webseite Hilfe

Deutsches Textarchiv

Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache

Das Deutsche Textarchiv stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit. Die Textauswahl erfolgte auf der Grundlage einer von Akademiemitgliedern erstellten und ausführlich kommentierten, umfangreichen Bibliographie. In Ergänzung wurden einschlägige Literaturgeschichten und (Fach-)Bibliographien ausgewertet. Aus der Gesamtliste der auf diesem Wege ermittelten Titel wurde von der DTA-Projektgruppe ein hinsichtlich der repräsentierten Textsorten und Disziplinen ausgewogenes Korpus zusammengestellt (weitere Informationen zur Textauswahl ...).

Um den historischen Sprachstand möglichst genau abzubilden, werden als Vorlage für die Digitalisierung in der Regel die Erstausgaben der Werke zugrunde gelegt. Das elektronische Volltextkorpus des DTA ist über das Internet frei zugänglich und dank seiner Aufbereitung durch (computer-)linguistische Methoden schreibweisentolerant über den gesamten jeweils verfügbaren Bestand durchsuchbar. Sämtliche Texte stehen zum Download zur Verfügung. *mehr ...*

Neuigkeiten aus dem Projekt

Dritter DTA-Workshop am 25. Oktober 2013

Zum Abschluss der zweiten Projektphase veranstaltet das DTA am 25. Oktober 2013 erneut einen Workshop zum Thema „Aufbau von Sprachressourcen am Beispiel des Deutschen Textarchivs". *mehr ...*

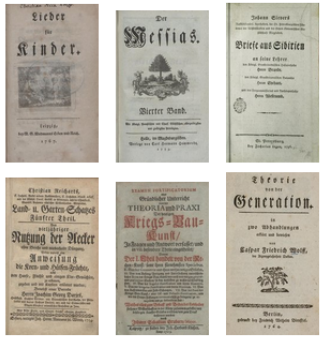
Das DTA auf der „TEI Conference and Members' Meeting 2013" in Rom

Vom 2. bis 5. Oktober findet die „TEI Conference and Members' Meeting 2013" in Rom statt. Das Deutsche Textarchiv ist mit einem Paper – Susanne Haaf und Alexander Geyken: „The Lifecycle of the DTA Base Format (DTABf)" – und einem Poster – Frank Wiegand: „TEI/XML Editing for Everyone's Needs" – vertreten. *mehr ...*

Das DTA in Zahlen

1013	Werke
345396	digitalisierte Seiten
80091342	fortlaufende Wortformen
563068924	Zeichen (Unicode)
542	weitere Werke in DTAQ

Neue Werke im DTA



alle Titel ...

Stöbern im DTA

- Textgattungen/-kategorien:
  - Belletristik
  - Gebrauchsliteratur
  - Wissenschaft
- Veröffentlichungszeit:
  - 17. Jahrhundert
  - 18. Jahrhundert
  - 19. Jahrhundert
  - 20. Jahrhundert

www.deutschestextarchiv.de

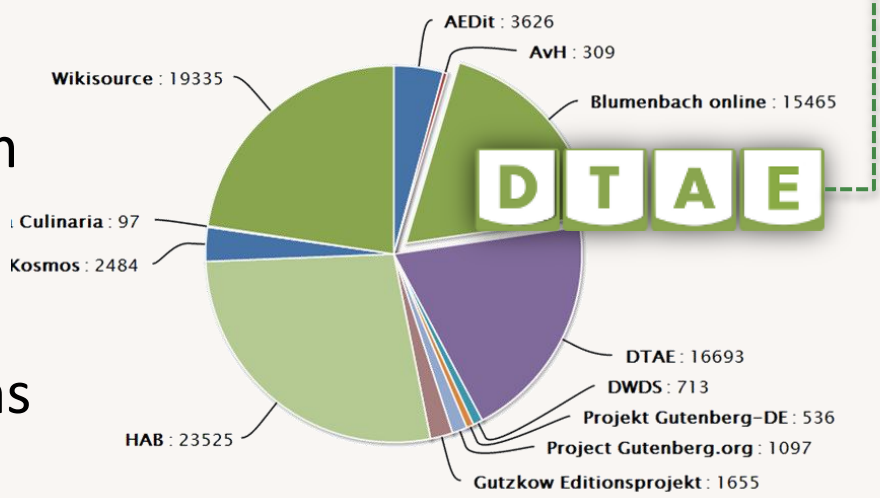
Deutsches Textarchiv (DTA): Überblick

- **Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache**
- Laufzeit: 2007–2014/15
- Förderung: DFG, Heimat: BBAW
- Partner in CLARIN-D

DTA

1 321	Werke
425 257	digitalisierte Seiten
99 779 178	fortlaufende Wortformen
696 929 675	Zeichen (Unicode)
854	weitere Werke in DTAQ

- disziplinen- und gattungs-übergreifender Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900
- DTA Kernkorpus+Erweiterungen: ~1,2 Mrd. Zeichen, ~200M Tokens



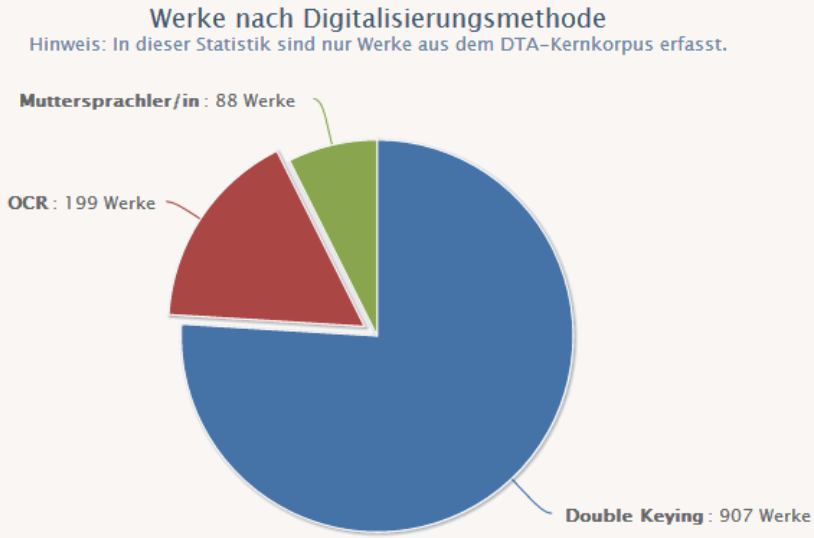
DTAE

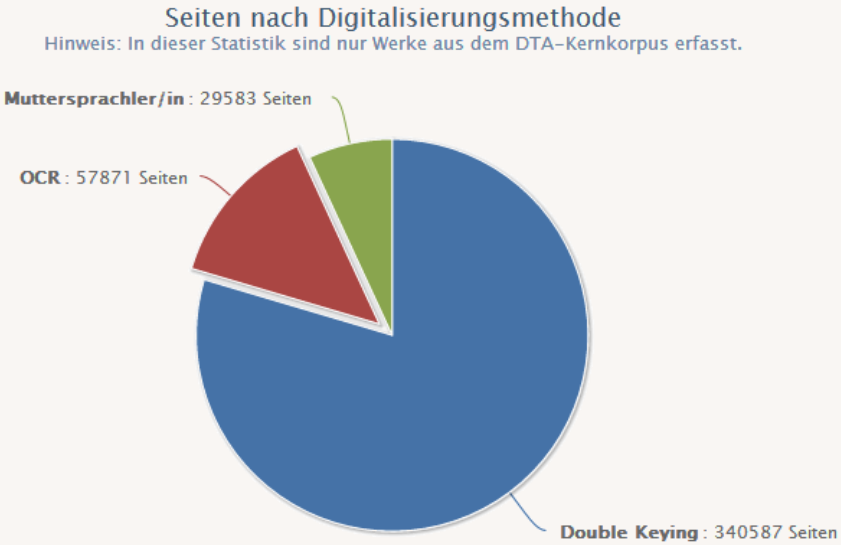
Texterfassung im DTA

- größtenteils manuelle Texterfassung per Double Keying
- formative und summative Qualitätssicherung

➢ **Zeichengenauigkeit >99,98%**

Cf. Studie Haaf/Wiegand/Geyken, jTEI 4 (2013), http://jtei.revues.org/739: **99.9909%**, d. h. 11 Transkriptionsfehler je 100000 Zeichen



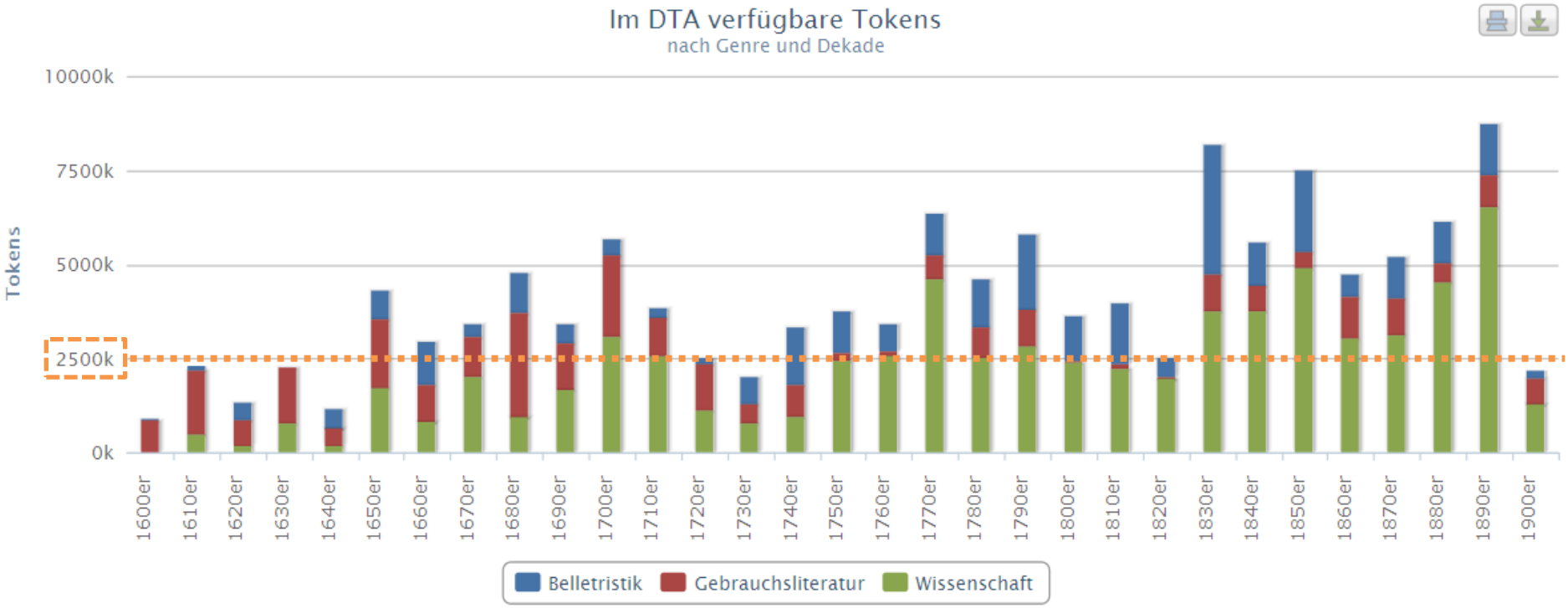


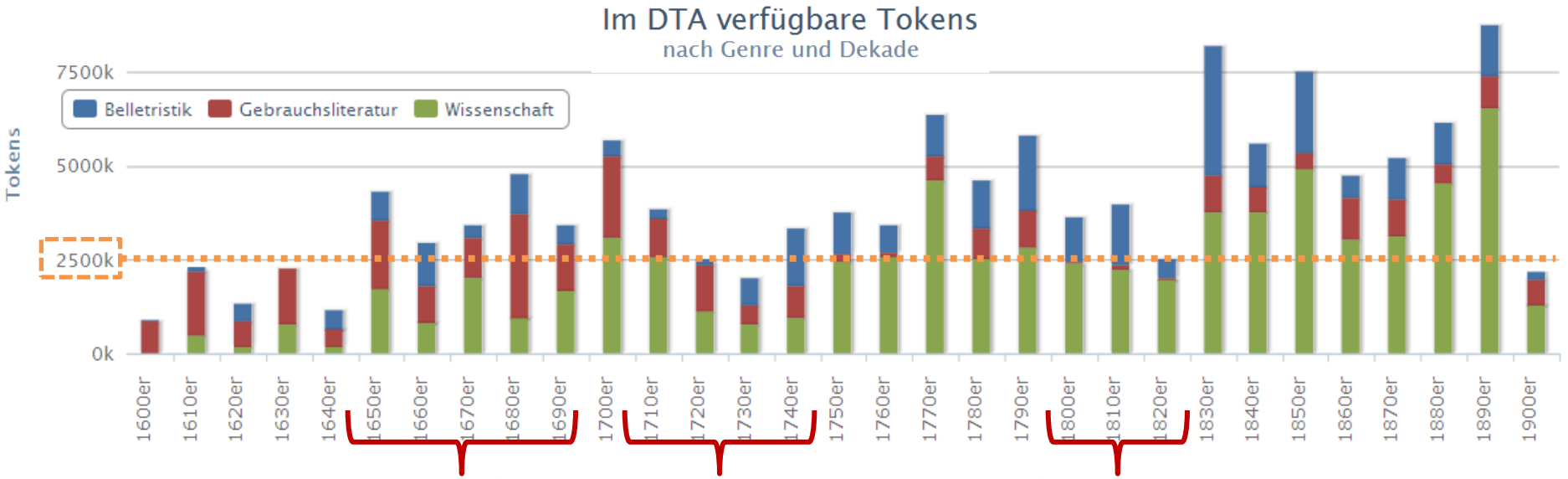
Verteilung DTA-Korpora nach Dekaden

Werke Gesamt:	**2060**
Belletristik:	520
Gebrauchsliteratur:	832
Wissenschaft:	708

Tokens Gesamt:	**132 055 789**
Belletristik:	29 162 714
Gebrauchsliteratur:	32 612 225
Wissenschaft:	70 280 850

Stand: 2014-07-04





*DTA Phase 3: **23683 S.***
(ca. 4,8 Mio. Tokens*)

Belletristik = 8978 S. (1,8 Mio. T.)
•lohenstein_cleopatra_1661 = 170 S.
•greiffenberg_sonnette_1662 = 478 S.
•weise_trauerspiel_1683 = 702 S.
•santa_judas02_1689+…04_1695 = 467 S., …

Gebrauchslit. = 6941 S. (1,4 Mio. T.)
•schupp_schriften_1663 = 1214 S.
•gerhardt_andachten_1667 = 528 S.
•spener_natur_1687 = 456 S.
•stieler_zeitungslust_1695 = 792 S., …

Wissenschaft = 7761 S. (1,6 Mio. T.)
•schottelius_arbeit_1663 = 1544 S.
•sturm_kunst_1670 = 462 S.
•dapper_amerika_1673 = 796 S.
•weirauch_dellaragione_1673 = 526 S., …

DTA Phase 2 in Arbeit/Phase 3:
30281 S. (ca. 6,1 Mio. Tokens*)

Belletristik = 5980 S. (1,2 Mio. Tokens)
•guenther_sammlung01_1724–…03_1727 = 1166 S.
•brockes_vergnuegen01_1724–…09_1748 = 2098 S.
•schnabel_fata01_1740+…04_1743 = 1203 S.
•gottsched_pietisterey_1736 = 176 S., …

Gebrauchslit. = 12940 S. (2,6 Mio. T.)
•wertheim_brieffsteller_1711 = 192 S.
•zinzendorf_christlichelehre_1725 = 317 S.
•bachstrom_schwimmen_1742 = 61 S.
•lucius_warnung_1747 = 464 S., …

Wissenschaft = 11361 S. (2,3 Mio. T.)
•decker_baumeister01_1711+…02_1713 = 147 S.
•mascov_geschichte01_1726+…02_1737 = 1224 S.
•schmeizel_staatswissenschafft_1732 = 302 S.
•baumgarten_aesthetica01_1750 = 412 S., …

*Hochrechnung,
Grundlage: 200 Tokens/S.

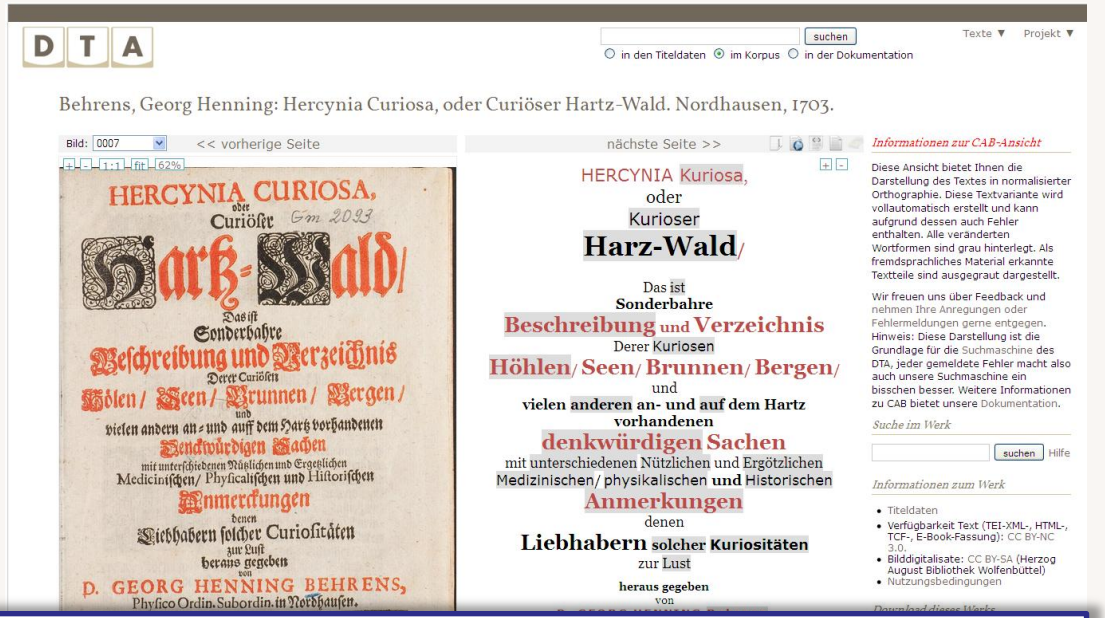
DTA Phase 3
5075 S. (ca. 1 Mio. Tokens*)

Belletristik = 2953 S. (0,6 Mio. Tokens)
•nettelbeck_lebensbeschreibung01_1821 +…02_1823+…03_1823 = 861 S.
•scheffner_leben_1823 = 582 S.
•waiblinger_phaeton01_1823+…02_1823 = 328 S.
•garve_briefe_1801 = 270
•wallenrodt_fritz01_1800+…02_1800 = 912 S.

Gebrauchslit. = 1230 S. (0,25 Mio. T.)
•schopenhauer_eyck01_1822 = 282 S.
•schopenhauer_eyck02_1822 = 220 S.
•fouque_frauen_1826 = 278 S.
•schmidlin_kinderindustrie_1821 = 157 S.
•schelle_spatziergaenge_1802 = 293

Wissenschaft = 892 S. (0,18 Mio. T.)
•chladni_akustik_1802 = 359 S.
•humboldt_geographie_1807 = 194 S.
•mueller_staatskunst03_1809 = 339 S.
+ 1820–1829: 34 Bde. Dinglers Polytechnisches Journal (Wissenschaft/Gebrauchsliteratur)

DTA: Key Features



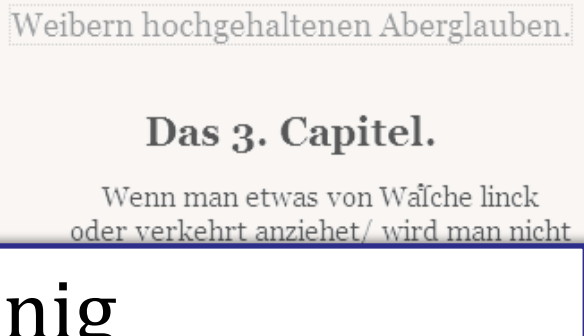
- Parallelansicht: facs|text
- Text in: TEI-XML, HTML, txt, CAB-View
- Metadaten: TEI-Header, CMDI, DC
- Linguistische Analyse (Standoff) → DDC
- Download: XML, HTML, plain text, TCF

➢ Download: DTA_core, DTA_core+DTE, alle Werke einer Kategorie, Zeitspanne etc.
www.deutschestextarchiv.de/download

Gliederung

- Textressourcen im WWW : Probleme aus der Nutzerperspektive
- Das Deutsche Textarchiv (DTA): Kurzvorstellung
- Vierter DTA-Workshop – Einführung in das Programm

Richtlinien zur Texterfassung



Matthias Boenig

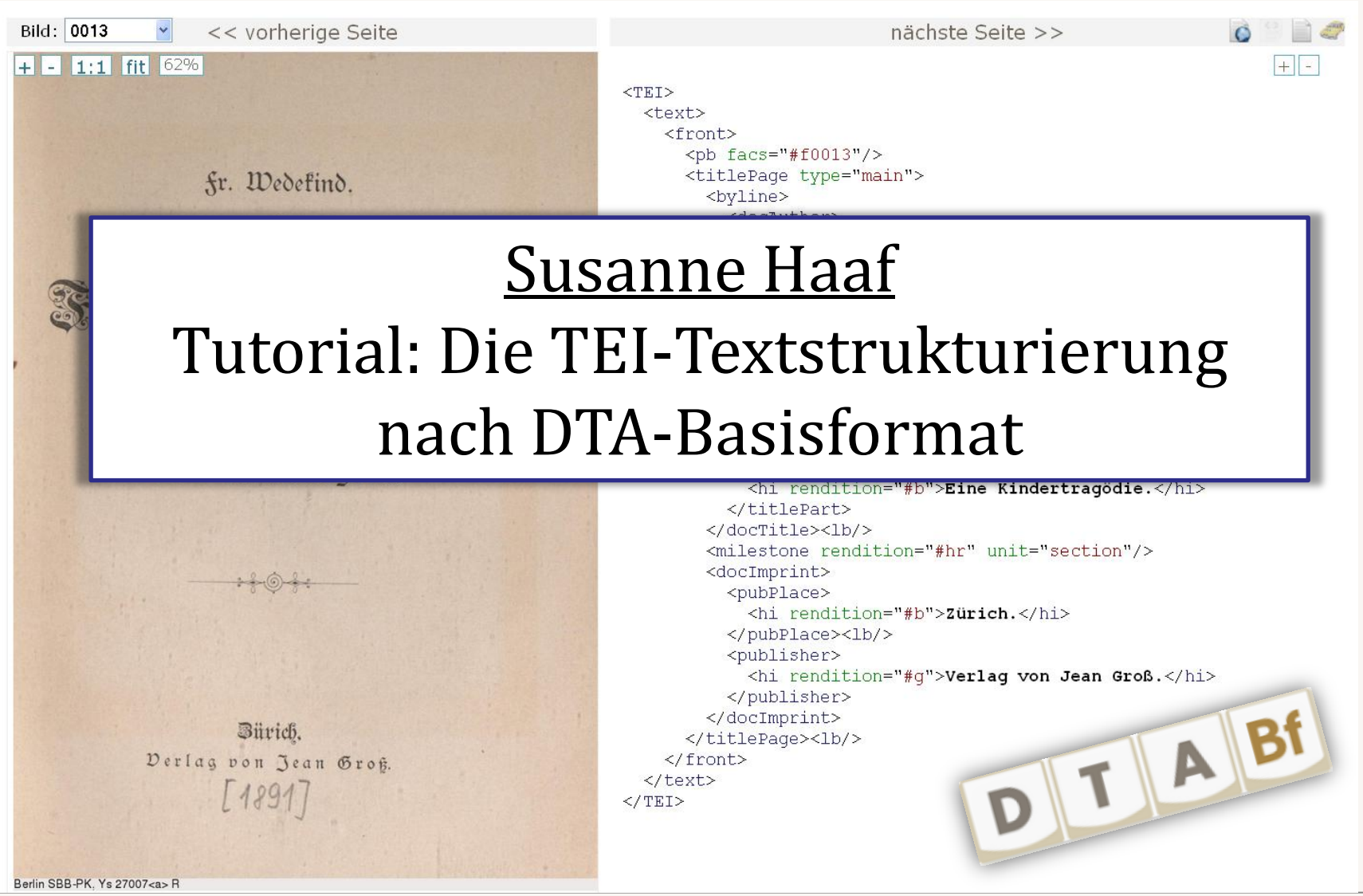
Tutorial: Ausgangsszenarien: Bild, Text, Annotation

- ze
- keine Modernisierungen; keine Normalisierungen
- „Ich laſſe mich nicht irre ſchreyn”

ſchreyn → ſchreyn ☺
ſchreyn → schreyn 😐
ſchreyn → schreien ☹

goethe_faust01_1808?p=293

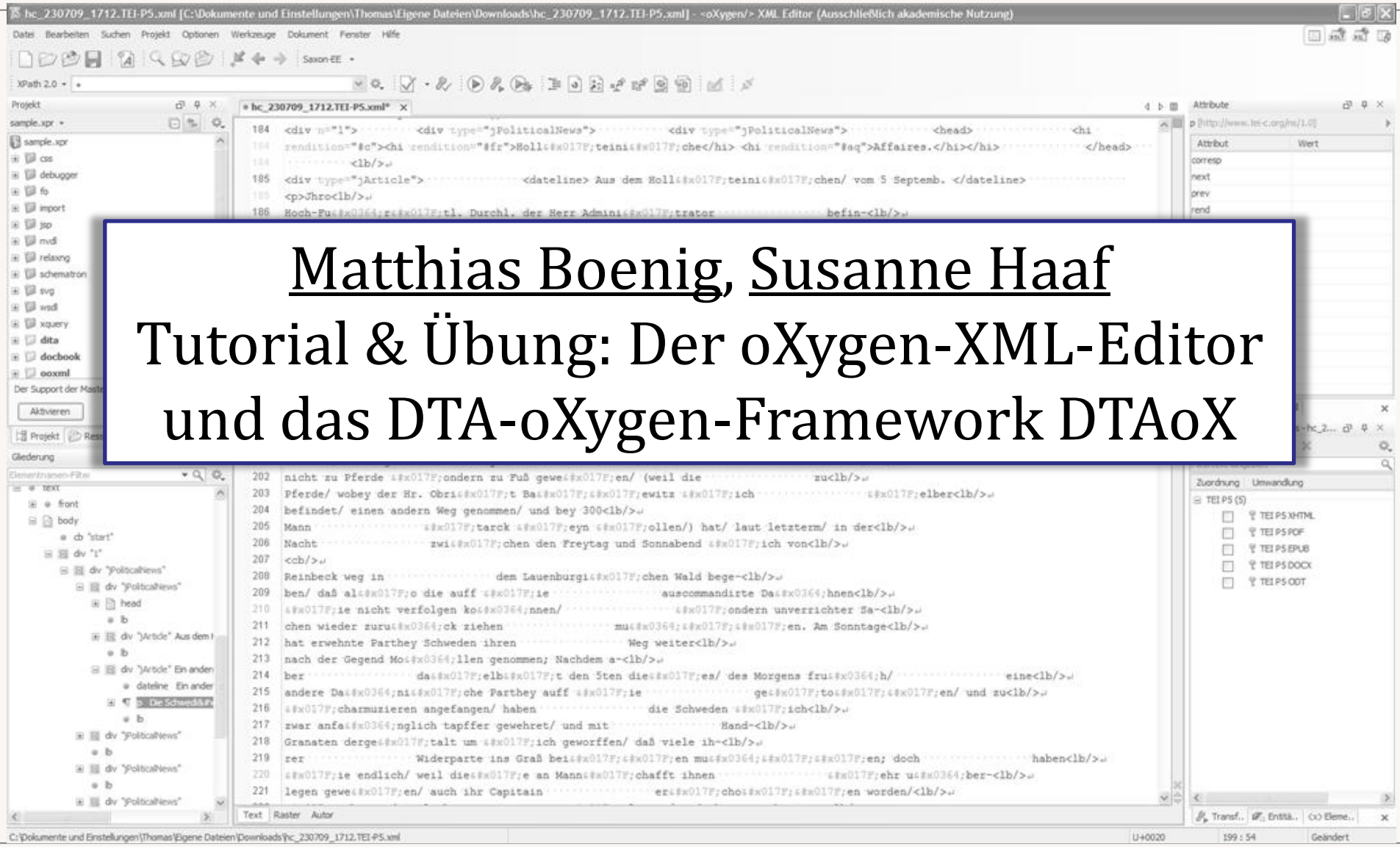
Richtlinien zur TEI/XML-Annotation



Susanne Haaf

Tutorial: Die TEI-Textstrukturierung nach DTA-Basisformat

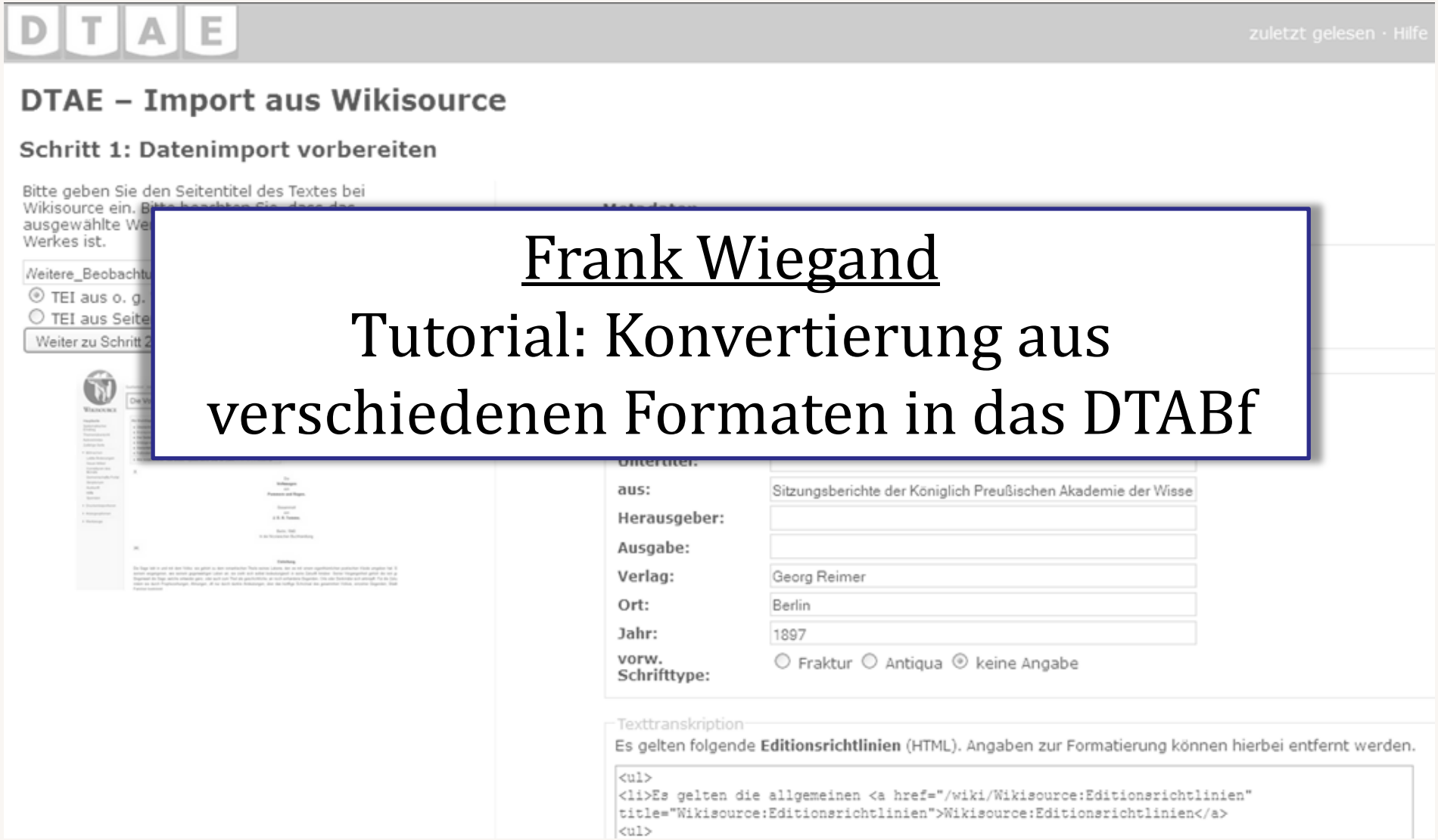
oXygen-XML-Editor



Matthias Boenig, Susanne Haaf

Tutorial & Übung: Der oXygen-XML-Editor und das DTA-oXygen-Framework DTAoX

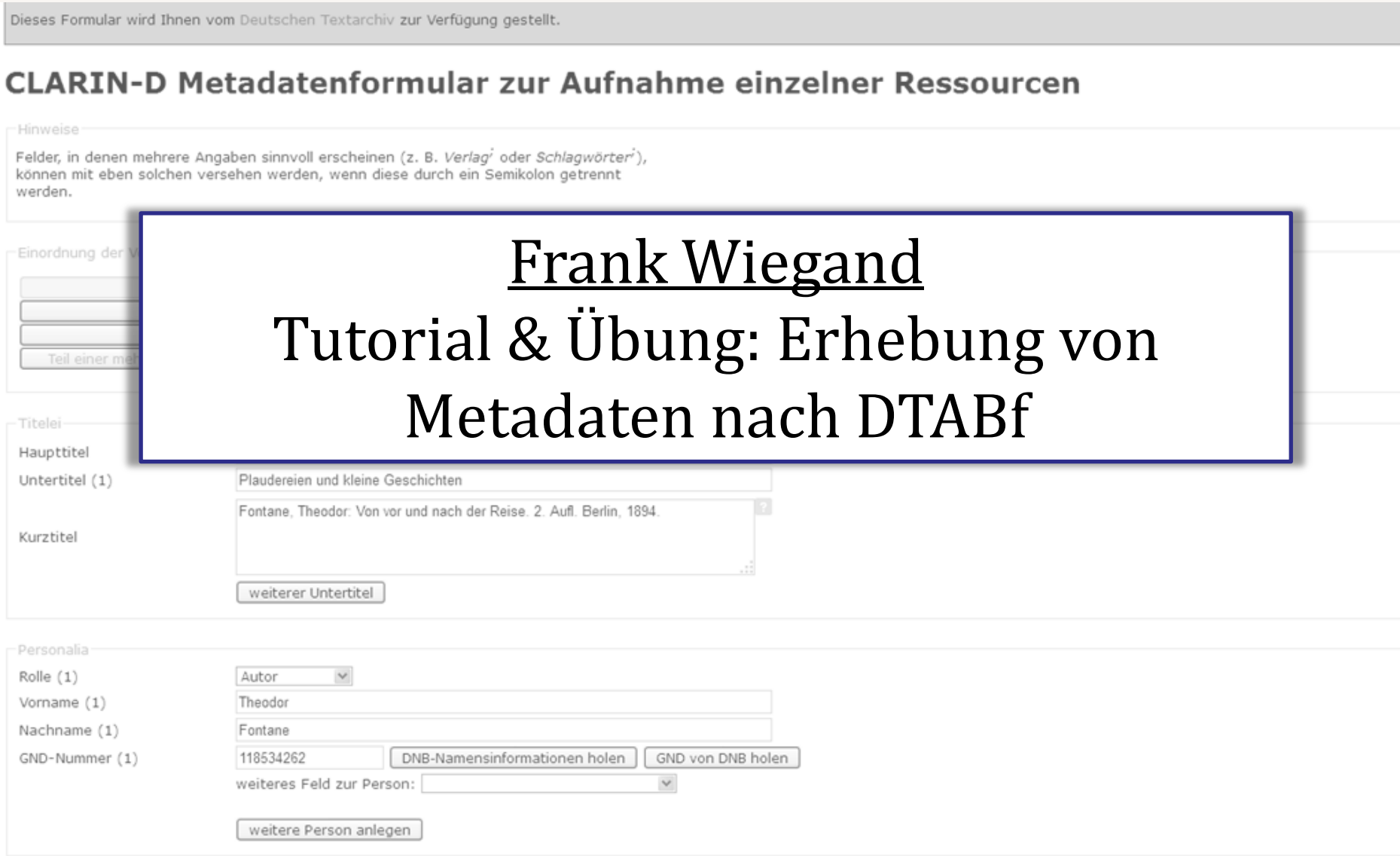
DTAE – Import aus Wikisource



Frank Wiegand

Tutorial: Konvertierung aus verschiedenen Formaten in das DTABf

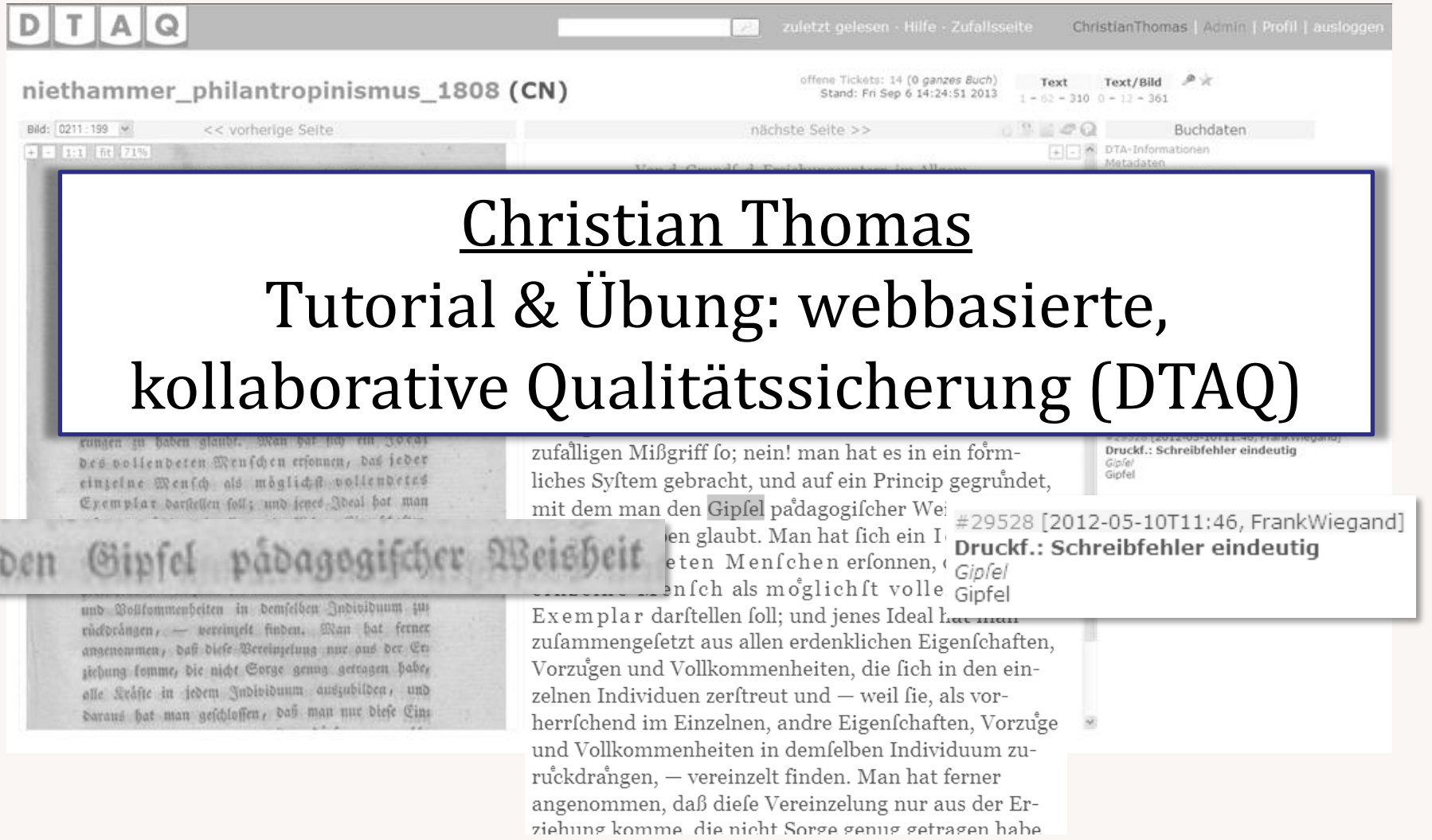
CLARIN-D Metadatenformular



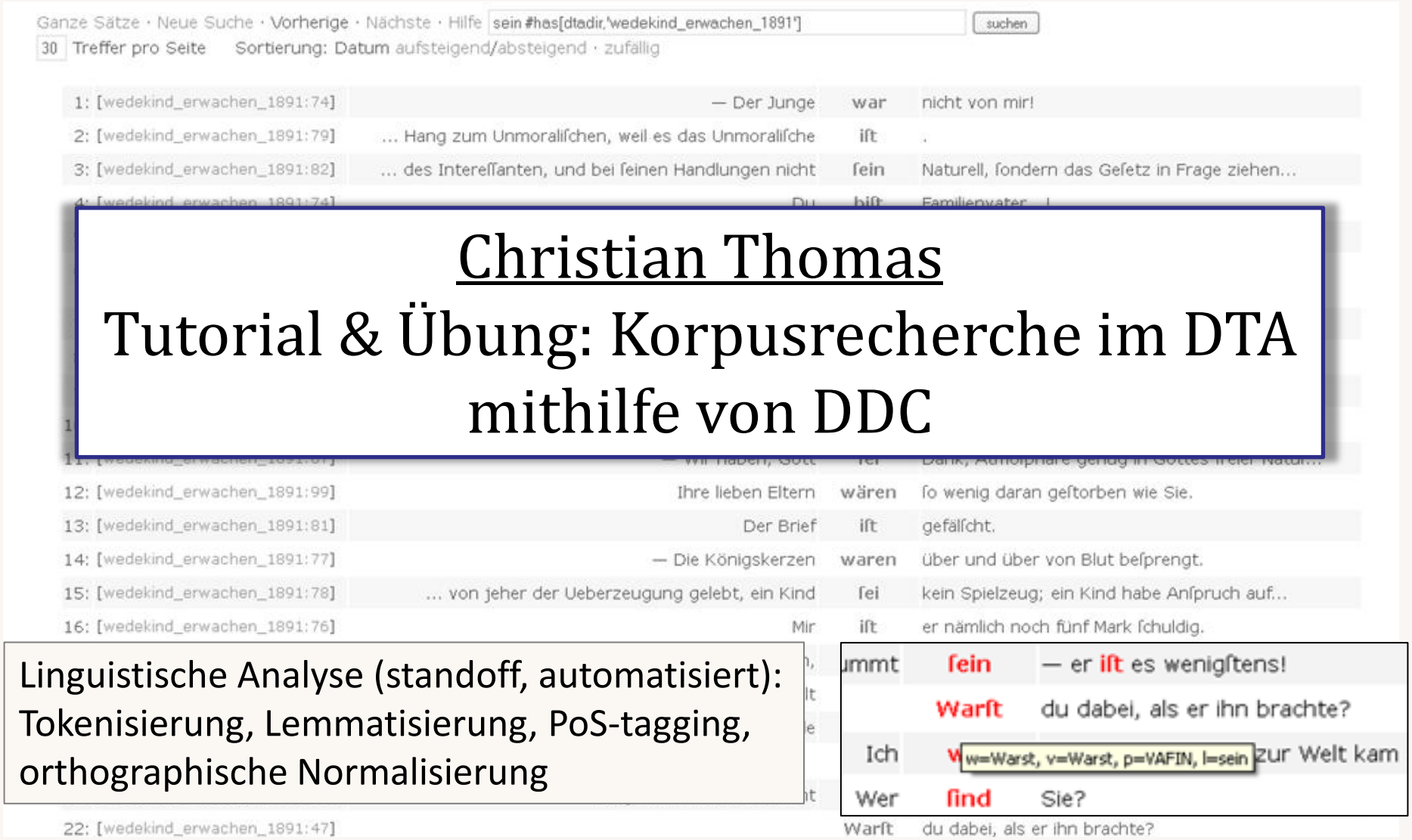
Frank Wiegand

Tutorial & Übung: Erhebung von Metadaten nach DTABf

Qualitätssicherung? DTAQ!



Linguistische Analyse / Suchmaschine DDC



Christian Thomas

Tutorial & Übung: Korpusrecherche im DTA mithilfe von DDC

Linguistische Analyse (standoff, automatisiert): Tokenisierung, Lemmatisierung, PoS-tagging, orthographische Normalisierung

CLARIN-D Federated Content Search

Susanne Haaf

Tutorial: Die DTA-Korpora in CLARIN-D

www.deutschestextarchiv.de

DTA

Matthias Boenig: Ausblick