Christian Thomas, Kay-Michael Würzner

Aufbereitung und Nachnutzung des „Deutschen Novellenschatzes“ im Deutschen Textarchiv

Team: Alexander Geyken, Matthias Boenig, Susanne Haaf, Bryan Jurish, Frank Wiegand

Berlin-Brandenburgische Akademie der Wissenschaften/Deutsches Textarchiv
dta@bbaw.de

CLARIN-D

berlin-brandenburgische AKADEMIE DER WISSENSCHAFTEN

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Das Deutsche Textarchiv (DTA)

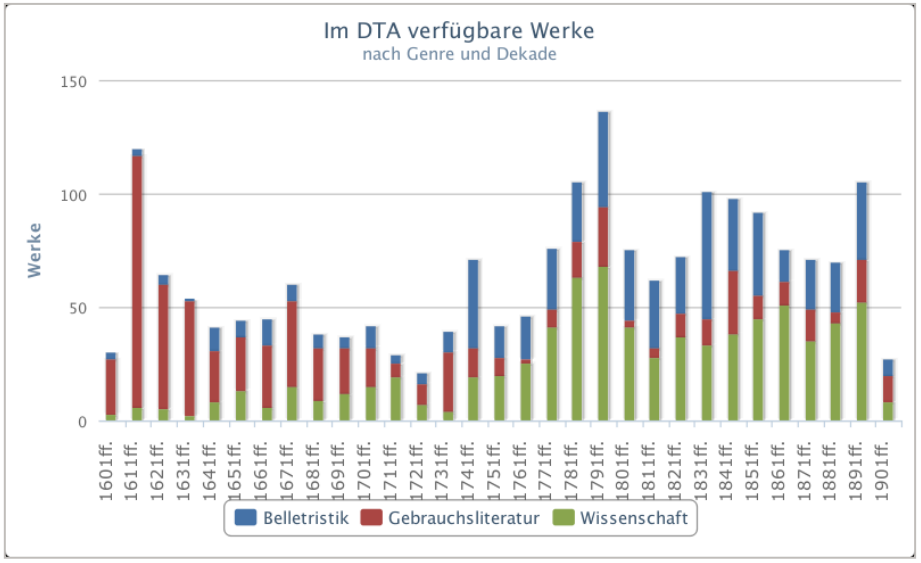
Deutsche Forschungsgemeinschaft

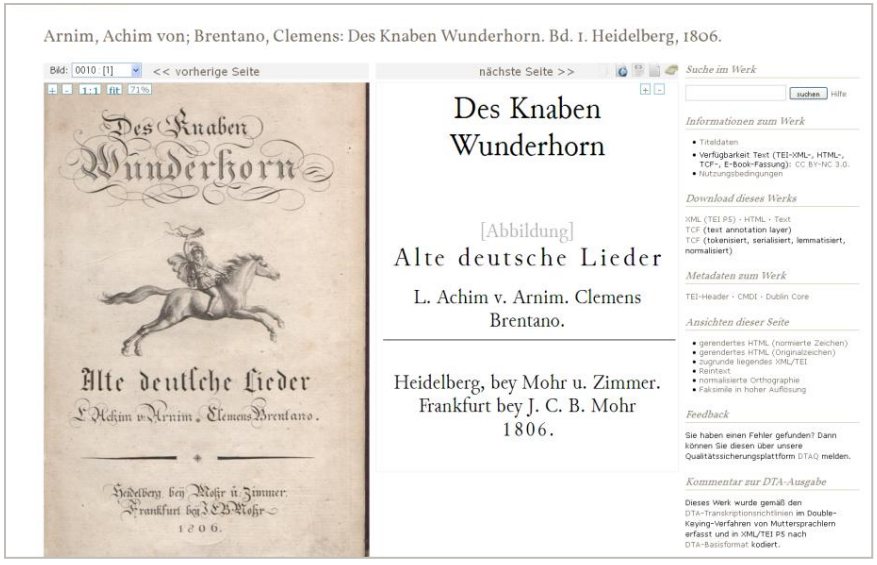
- Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache
- deutschsprachige Werke des 17. bis 19. Jhs.
- diverse Textsorten und wissenschaftliche Disziplinen

www.deutschestextarchiv.de

Das DTA in Zahlen Stand: 8.5.2017

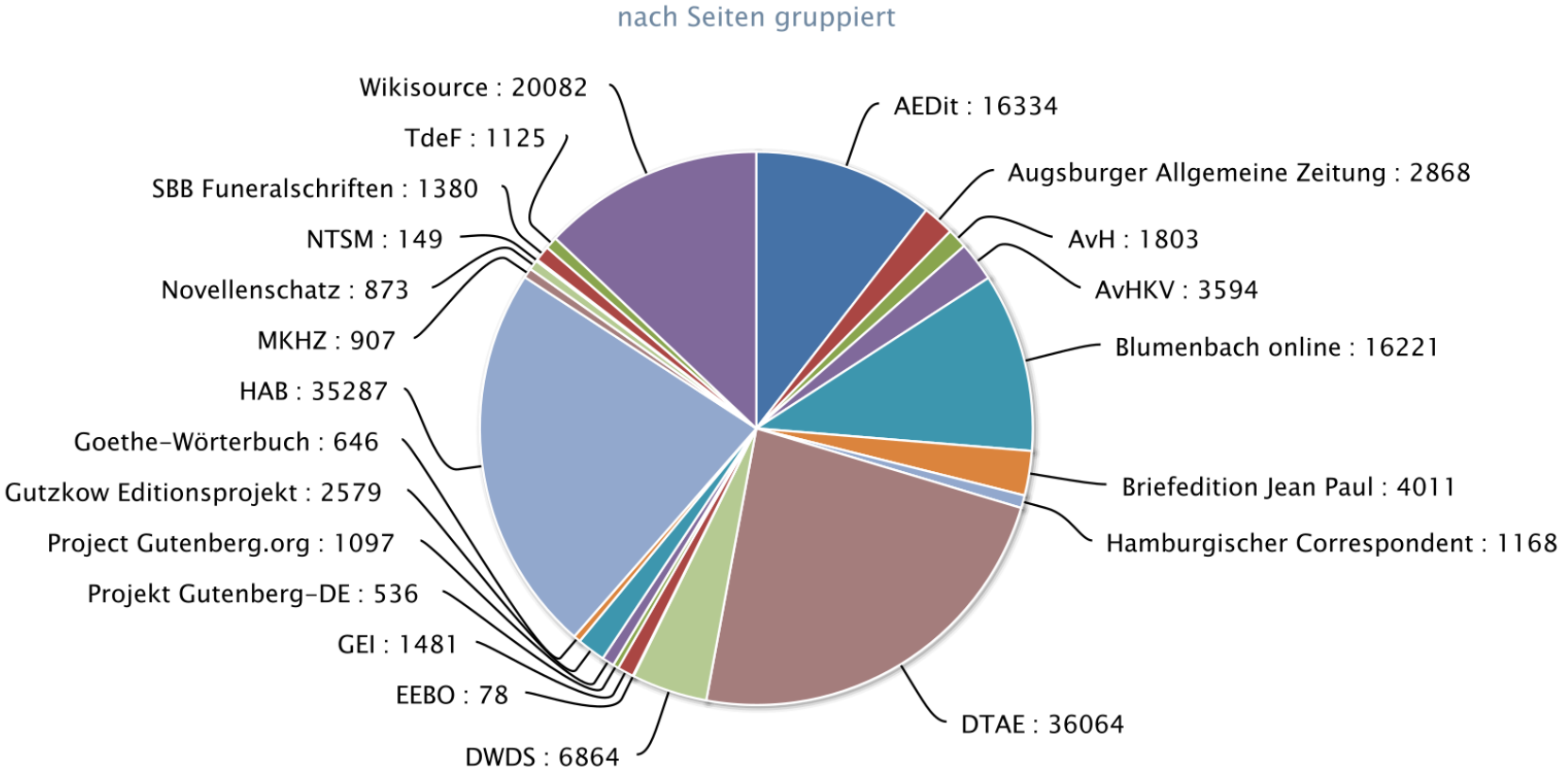
2 631	Werke
612 549	digitalisierte Seiten
146 727 000	fortlaufende Wortformen
1 025 731 190	Zeichen (Unicode)
900	weitere Werke in DTAQ





DTA-Erweiterungen

- weitere Werke aus externen Quellen
- Anpassung an DTA-Richtlinien → Homogenität, Interoperabilität
- Vernetzung von ursprünglich separaten Projekten
- > 30 Projekte, 1.862 Dokumente, ca. 165.000 Seiten



DTA-Erweiterungen

- weitere Werke aus externen Quellen
- Anpassung an DTA-Richtlinien → Homogenität, Interoperabilität
- Vernetzung von ursprünglich separaten Projekten
- > 30 Projekte, 1.862 Dokumente, ca. 165.000 Seiten, z.B.:
  - Wikisource (Diverse, ≈ 20.000 Seiten)
  - Blumenbach-online (dt. Monographien B.s, 1779–1823 ≈ 16.200 S.)
  - AEDit (Funeralschriften, 1562–1742, ≈ 16.300 S.)
  - ePoetics (Poetiken, 1804–1969, ≈ 10.000 S.)
  - Humboldts Kosmos-Vorträge (1827/28, hs. Nachschriften, ≈ 3.600 S.)
  - Augsburger Allgemeine Zeitung (1840, ≈ 3.000 S.),
    Hamburgischer Correspondent (1712–1851, ≈ 1.200 S.)
- plus zwei vollständige historische Journale:
  - Polytechnisches Journal (1821–1931), 375 Bde., ≈ 205.000 S.
  - Die Grenzboten (1841–1922), 300 Bde., ≈ 180.000 S.

Das Deutsche Textarchiv (DTA)

- Korpusgröße (inkl. DTAQ, ohne Dingler u. Grenzboten):

Werke	3 531
Seiten	702 093
Tokens	ca. 176 Mio.
Zeichen	ca. 1,224 Mrd.

Stand: 8.5.2017

- strukturelles/semantisches Tagging: TEI nach DTA-Basisformat
- Metadaten: TEI-Header, Dublin Core, CMDI
- Download und Harvesting der Daten
  - verschiedene Formate, freie Lizenzen
- Linguistische Analyse:
  - Tokens, Lemmata, Wortart (PoS), orthographische Normierung, distributionelle Semantik
  - Indizierung mit der Suchmaschine DDC

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Grundstruktur jedes TEI-Dokuments

Metadaten
zum digitalen Objekt, zur Vorlage, zum Encoding, …

Objektdaten
annotierte Transkription der Vorlage

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>...</fileDesc>
    <encodingDesc>...</encodingDesc>
    <profileDesc>...</profileDesc>
  </teiHeader>
  <text>
    <front>
      <p>Elemente vor Beginn des Buchtextes</p>
    </front>
    <body>
      <p>Textkörper</p>
    </body>
    <back>
      <p>Elemente nach Abschluss des Buchtextes</p>
    </back>
  </text>
</TEI>
```

Das DTA-Basisformat (DTABf)

- TEI-Format für die Auszeichnung von historischen Texten in Korpora
- Schema & Dokumentation
- Ziel: Ambiguität vermeiden, Interoperabilität gewährleisten
- mittlerweile DFG-Empfehlung

Neu: DTABf für Manuskripte

DTABf

EINFÜHRUNG | METADATEN | TRANSKRIPTION | FORMAL | INHALTLICH | SPEZIAL | ÜBERSICHTEN | KONTAKT

DTA-Basisformat – Auszeichnung von Manuskripten

Inhaltsverzeichnis dieses Dokuments

1 Allgemeines
- 1.1 Geltungsbereich der vorliegenden Richtlinien
- 1.2 Terminologische Konventionen der vorliegenden Richtlinien

2 Besonderheiten der Kapitelstruktur
- 2.1 Kapitel
- 2.2 Titel

3 Sekundäre Textänderungen des Autors und späterer Bearbeiter
- 3.1 Hinzufügungen und Tilgungen
- 3.2 Substitution
- 3.3 Einweisung einer Hinzufügung in den Text
- 3.4 Anmerkungen des Autors/späterer Bearbeiter

4 Platzhalter des Autors

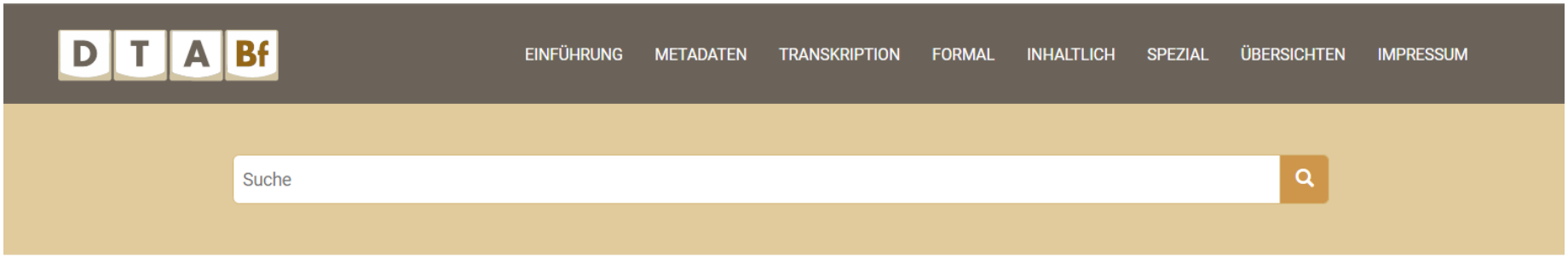
5 Inventar des Editors der digitalen Ausgabe
- 5.1 Verantwortlichkeiten
- 5.2 Unsichere Lesarten

6 Seiten- und Blattnummerierung

7 Wiedergabe von Bibliotheksstempeln

8 Besonderheiten bei den Metadaten

Das DTA-Basisformat (DTABf)



- DTABf-konforme XML-Vorlagendatei
- RNG-Schema, Schematron-Regelsatz
- Metadatenformular
- Richtlinien zur Transkription

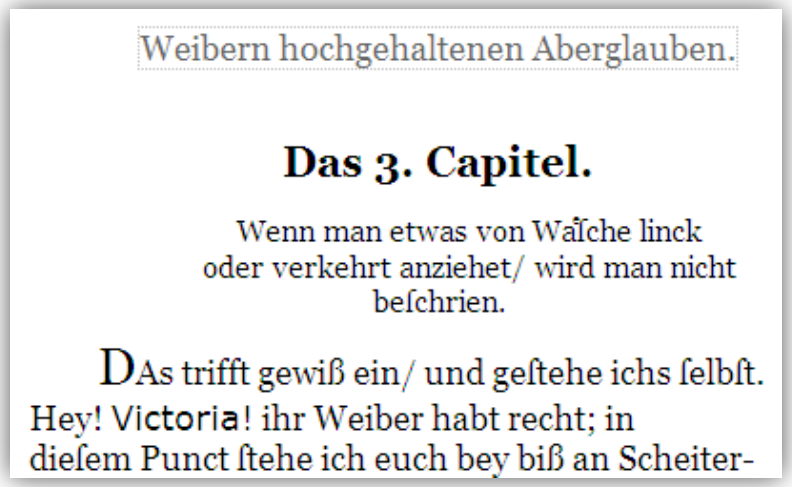
DTA-Richtlinien zur Transkription

Weibern hochgehaltenen Aberglauben. 5

Das 3. Capitel.

Wenn man etwas von Wäſche linck oder verkehrt anzieher/ wird man nicht beſchrien.

DAs trifft gewiß ein/ und geſtehe ichs ſelbſt. Hey! Victoria! ihr Weiber habt recht; in dieſem Punct ſtehe ich euch bey biß an Scheiter-


Weibern hochgehaltenen Aberglauben.

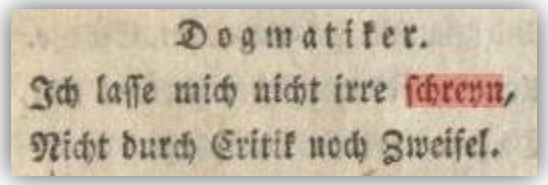
Das 3. Capitel.

Wenn man etwas von Waſche linck oder verkehrt anziehet/ wird man nicht beſchrien.

DAs trifft gewiß ein/ und geſtehe ichs ſelbſt. Hey! Victoria! ihr Weiber habt recht; in dieſem Punct ſtehe ich euch bey biß an Scheiter-

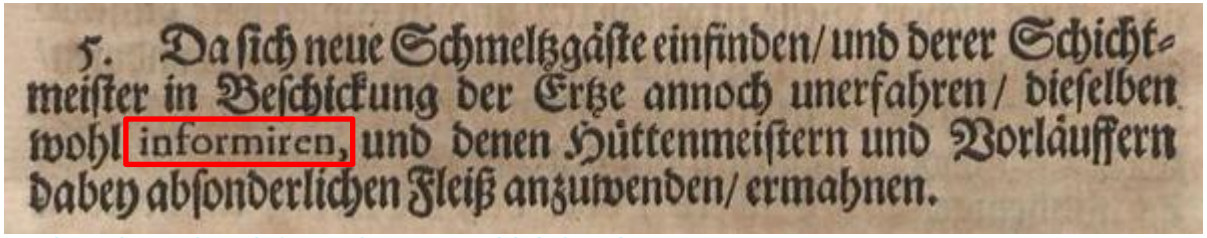
- seiten-, zeilen- und zeichengetreu nach Vorlage
- Strukturierung nach Vorlage (ggf. implizit → explizit)
- keine Modernisierungen; keine Normalisierungen:
  „Ich laſſe mich nicht irre ſchreyn”
  - ſchreyn → ſchreyn ☺
  - ſchreyn → schreyn 😐
  - ſchreyn → schreien ☹

goethe_faust01_1808/293

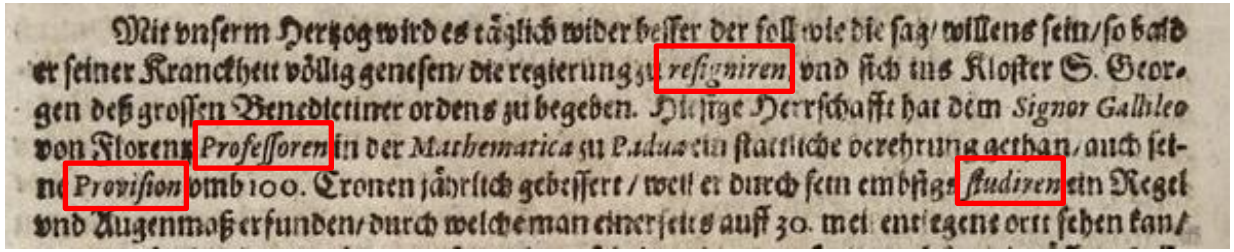

Dogmatiker.

Ich laſſe mich nicht irre ſchreyn,
Nicht durch Critik noch Zweifel.

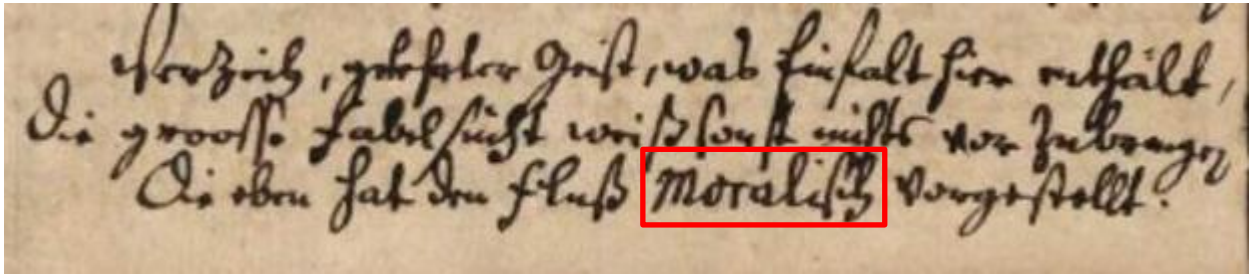
Das DTA-Basisformat (DTABf)


5. Da sich neue Schmeltzgäste einfinden/ und derer Schicht-
meister in Beschickung der Ertze annoch unerfahren/ dieselben
wohl informiren, und denen Hüttenmeistern und Vorläuffern
dabey absonderlichen Fleiß anzuwenden/ ermahnen.

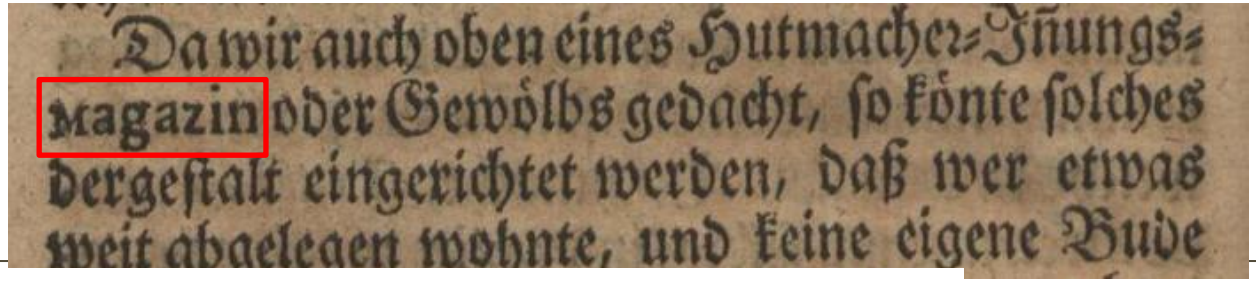
schoenberg_berginformation_1693?hl=informirenp=103


Mit vnserm Hertzog wird es täglich wider besser der soll wie die sag/ willens sein/ so bald
er seiner Kranckheit völlig genesen/ die regierung zu *resigniren* vnd sich ins Kloster S. Geor-
gen deß grossen Benedictiner ordens zu begeben. Hiesige Herrschafft hat dem *Signor Galileo*
von Florenz *Professoren* in der *Mathematica* zu *Padua* ein stattliche verehrung gethan/ auch sei-
ne *Provision* vmb 100. Cronen jährlich gebessert/ weil er durch sein embsiges *studiren* ein Regel
vnd Augenmaß erfunden/ durch welche man einerseits auff 30. meil entlegene ort sehen kan/

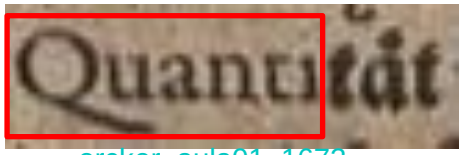
anonym_relation_1609/?hl=studiren&p=164


Herzich, gehaftes Gest, wab fiefalt sin rulfalt,
Di groose Fabelfust wirstbart mist der Jubrang?
Di eben sat den fluß moralisch vorgestellt?

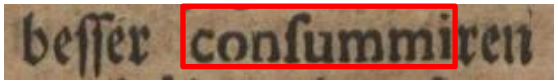
zeller_chronik01_1700/?hl=moralisch&p=59


Da wir auch oben eines Hutmacher-Jnungs-
Magazin oder Gewölbs gedacht, so könte solches
dergestalt eingerichtet werden, daß wer etwas
weit abgelegen wohnte, und keine eigene Bude

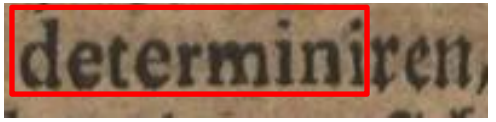
marperger_hutmacher_1719/?hl=con%C5%BFummiren&p=100


Quantität

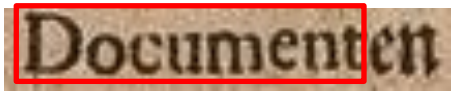
ercker_aula01_1672


besser consummiren

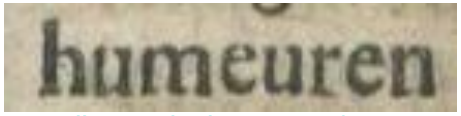
marperger_hutmacher_1719/?p=97


determiniren,

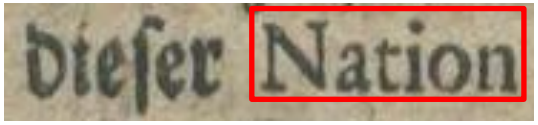
marperger_hutmacher_1719/?p=97


Documenten

fleming_jaeger01_1719/?p=599


humeuren

mueller_ostiacken_1726/?p=55


dieser Nation

mueller_ostiacken_1726/?p=55

Das DTA-Basisformat (DTABf)

Ziel: Vermeidung von Ambiguitäten
→ Interoperabilität

TEI

```
<rs type="propNounPersName">Sokrates</rs>,
<name type="person">Platon</name> und
<persName>Aristoteles</persName> waren Philosophen.
```

DTABf

```
<persName>Sokrates</persName>,
<persName>Platon</persName> und
<persName>Aristoteles</persName> waren Philosophen.
```

Beispiel: Novellenschatz / Einleitung

DTAQ | zuletzt gelesen · Hilfe · Zufallsseite | ChristianThomas | Admin | Profil | ausloggen

alexis_sacken_1910 (Novellenschatz)

offene Tickets: 0 (0 *ganzes Buch*)
Stand: 2017-05-05 17:17:09

Text	Text/Bild
0 – 0 – 110	0 – 0 – 110

Bild: 0005 ▾ << vorherige Seite

```
<TEI xmlns="http://www.tei-c.org/ns/1.0"><text><front><pb facs="#f0005"/>
<div type="preface">              <p><persName ref="http://d-nb.info/gnd/118648071">G. Wilhelm Heinrich Häring</persName>, geboren zu Breslau den 29. Juni 1798, aus einer Refugié-Familie der Bretagne stammend, erlebte als Knabe 1806 die Belagerung Breslau's,              wurde nach dem Tode des Vaters nach Berlin verpflanzt, wo er das Werder'sche Gymnasium besuchte, machte als Freiwilliger 1815 den Feldzug und die Belagerungen der              Ardennenfestungen mit, studirte seit 1817 in Berlin und Breslau die Rechte und trat in              den Staatsdienst, aus welchem er aber schon als Kammergerichtsreferendarius zur              Literatur (unter dem Namen <persName ref="http://d-nb.info/gnd/118648071">Wilibald Alexis</persName>) überging.</p><lb/>
<p>Nachdem er durch den unter <persName ref="http://d-nb.info/gnd/118612409">Walter Scott</persName>'s Maske geschriebenen, wirklich ebenbürtigen und              überall als echt aufgenommenen, von <persName ref="http://d-nb.info/gnd/118612409">Scott</persName> selbst für die kühnste Mystification des              Jahrhunderts erklärten Roman „Walladmor“ (welchem „Schloß Avalon“ folgte) seinen Ruf weithin begründet, verdiente er sich durch die selbstständigen großen Romane aus der              Geschichte Brandenburg-Preußens, „Cabanis“, „Roland von Berlin“, „Der letzte Waldemar“, „Dorothee“, „Die Hosen des Herrn von Bredow“ ꝛc. ꝛc., den Namen des „preußischen <persName ref="http://d-nb.info/gnd/118612409">Walter              Scott</persName>“. Allgemeine Verbreitung erlangte seine mit <persName ref="http://d-nb.info/gnd/119209349">Hitzig</persName> unternommene Sammlung von              Criminalfällen: „Der neue Pitaval“.</p><lb/>
<p>Am 16. December 1871 nahm ihn der Tod aus einem durch sechzehn Jahre in voller              Heiterkeit des Gemüthes erduldeten Leidenszustande hinweg.</p><lb/>
<p>Aus seinen zahlreichen Novellen, worin er die <persName ref="http://d-nb.info/gnd/12989432X">Tieck</persName>'sche Schule allmählich mit einer ihm              durchaus eigenthümlichen Kunstweise vertauschte, eine einzelne auszuwählen, ist keine leichte Aufgabe. <persName ref="http://d-nb.info/gnd/118648071">Häring</persName> ist nicht eigentlich Novellist im strengen<lb/>
</p></div></front></text></TEI>
```

nächste Seite >>

G. Wilhelm Heinrich Häring, geboren zu Breslau den 29. Juni 1798, aus einer Refugié-Familie der Bretagne stammend, erlebte als Knabe 1806 die Belagerung Breslau's, wurde nach dem Tode des Vaters nach Berlin verpflanzt, wo er das Werder'sche Gymnasium besuchte, machte als Freiwilliger 1815 den Feldzug und die Belagerungen der Ardennenfestungen mit, studirte seit 1817 in Berlin und Breslau die Rechte und trat in den Staatsdienst, aus welchem er aber schon als Kammergerichtsreferendarius zur Literatur (unter dem Namen Wilibald Alexis) überging.

Nachdem er durch den unter Walter Scott's Maske geschriebenen, wirklich ebenbürtigen und überall als echt aufgenommenen, von Scott selbst für die kühnste Mystification des Jahrhunderts erklärten Roman „Walladmor“ (welchem „Schloß Avalon“ folgte) seinen Ruf weithin begründet, verdiente er sich durch die selbstständigen großen Romane aus der Geschichte Brandenburg-Preußens, „Cabanis“, „Roland von Berlin“, „Der letzte Waldemar“, „Dorothee“, „Die Hosen des Herrn von Bredow“ ꝛc. ꝛc., den Namen des „preußischen Walter Scott“. Allgemeine Verbreitung erlangte seine mit Hitzig unternommene Sammlung von Criminalfällen: „Der neue Pitaval“.

Am 16. December 1871 nahm ihn der Tod aus einem durch sechzehn Jahre in voller Heiterkeit des Gemüthes erduldeten Leidenszustande hinweg.

Aus seinen zahlreichen Novellen, worin er die Tieck'sche Schule allmählich mit einer ihm durchaus eigenthümlichen Kunstweise vertauschte, eine einzelne auszuwählen, ist keine leichte Aufgabe. Häring ist nicht eigentlich Novellist im strengen

Beispiel: Personen Kosmos-Vorträge

```
1987  <p>Trotz dem bleiben manche Erscheinungen: das Nordlicht,<lb/>
1988  <add place="left">Wolkenbildungen, </
1989  ein unerklärliches Dunkel gehült. <pe
1990  ····<persName resp="#CT" ref="http://
1991  ····<persName resp="#SB" ref="http://
1992  ····<persName resp="#CT" ref="http://
1993  ····&amp;amp; <persName resp="#SB" ref="h
1994  ···haben in diesem Fache<lb/>ausgeze
1995  <p>In der Optik und Astronomie sind d
1996  <persName resp="#SB" ref="http://d-nb
1997  ····<persName resp="#SB" ref="http://
1998  ····<persName resp="#SB" ref="http://
1999  se gewesen: man fand am Ende unseres
```

Name (Ansetzungsform der GND)	Vorkommen	Vorlageform(en)	Normdaten
Barberini, Francesco (1597–1679)	1 + 1 (2)	Bambini ↗, ↗	http://d-nb.info/gnd/100013376
Bory de Saint-Vincent, Jean-Baptiste Geneviève Marcellin (1778–1846)	1 + 3 + 1 (5)	Bory de St Vincent ↗, ↗, ↗ Bory St. Vincent ↗ Bory de St. Vincent ↗	http://d-nb.info/gnd/100050395
Brandes, Heinrich Wilhelm (1777–1834)	2 + 2 + 3 + 1 + 3 + 2 + 1 (14)	Brandes ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗ Brandis ↗	http://d-nb.info/gnd/100055427
Crawfurd, John (1783–1868)	1 (1)	Crawfurd ↗	http://d-nb.info/gnd/100094546
Erman, Adolf (1806–1877)	1 + 1 + 2 + 1 + 1 + 1 (7)	Erman ↗, ↗, ↗, ↗, ↗, ↗, ↗	http://d-nb.info/gnd/100123848
Freytag, Georg Wilhelm (1788–1861)	1 + 1 + 1 + 1 (4)	Freitag ↗, ↗, ↗, ↗	http://d-nb.info/gnd/100126707
	1 + 1 (2)	Fischer ↗, ↗	http://d-nb.info/gnd/100132936
	1 + 1 (2)	Jean Sanac ↗ Jean Lou ↗	http://d-nb.info/gnd/100141498
	1 (1)	Legentil ↗	http://d-nb.info/gnd/100188265
	1 + 1 + 2 + 1 + 1 + 1 (7)	Lexel ↗, ↗, ↗, ↗, ↗ Lexly ↗ Leccie ↗	http://d-nb.info/gnd/100189180
	4 + 3 + 2 + 3 + 9 + 3 + 2 (29)	Lichtenstein ↗, ↗ Lichtenltein ↗, ↗, ↗, ↗, ↗, ↗	http://d-nb.info/gnd/100190014
	1 + 1 (2)	Kries ↗ Gries ↗	http://d-nb.info/gnd/100203604
	1 + 1 + 1 + 2 + 1 + 1 (7)	Morin ↗, ↗, ↗, ↗, ↗, ↗, ↗	http://d-nb.info/gnd/100238750
	5 + 4 + 1 + 7 + 3 + 1 + 2 (23)	Parrow ↗, ↗ Parrer ↗, ↗ Parot ↗, ↗ Parrot ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗, ↗ Parrol ↗, ↗ Sarrot ↗ Paroot ↗	http://d-nb.info/gnd/100240445
	3 + 1 + 1 + 2 + 6 + 6 + 3 (23)	Abel-Rémusat ↗, ↗ Abel Rémusat ↗, ↗, ↗, ↗ Abel Remusat ↗, ↗, ↗	http://d-nb.info/gnd/100318371

Personenverzeichnis „Hidden Kosmos“

Alle Dokumente (Stand: 2017-04-06 11:37:20)
Auszeichnungen mit <persName> insgesamt: 8820
<persName>s mit @ref: 8590
<persName>s ohne @ref: 230
unique Personen: 749 (+ 230 ohne @ref)

hufeland_privatbesitz_1829
Auszeichnungen mit <persName> insgesamt: 505
<persName>s mit @ref: 500
<persName>s ohne @ref: 5
unique Personen: 280 (+ 5 ohne @ref)

nn_msgermqu2124_1827
Auszeichnungen mit <persName> insgesamt: 478
<persName>s mit @ref: 474
<persName>s ohne @ref: 4
unique Personen: 280 (+ 4 ohne @ref)

nn_msgermqu2345_1827
Auszeichnungen mit <persName> insgesamt: 1172
<persName>s mit @ref: 1145
<persName>s ohne @ref: 27
unique Personen: 445 (+ 27 ohne @ref)

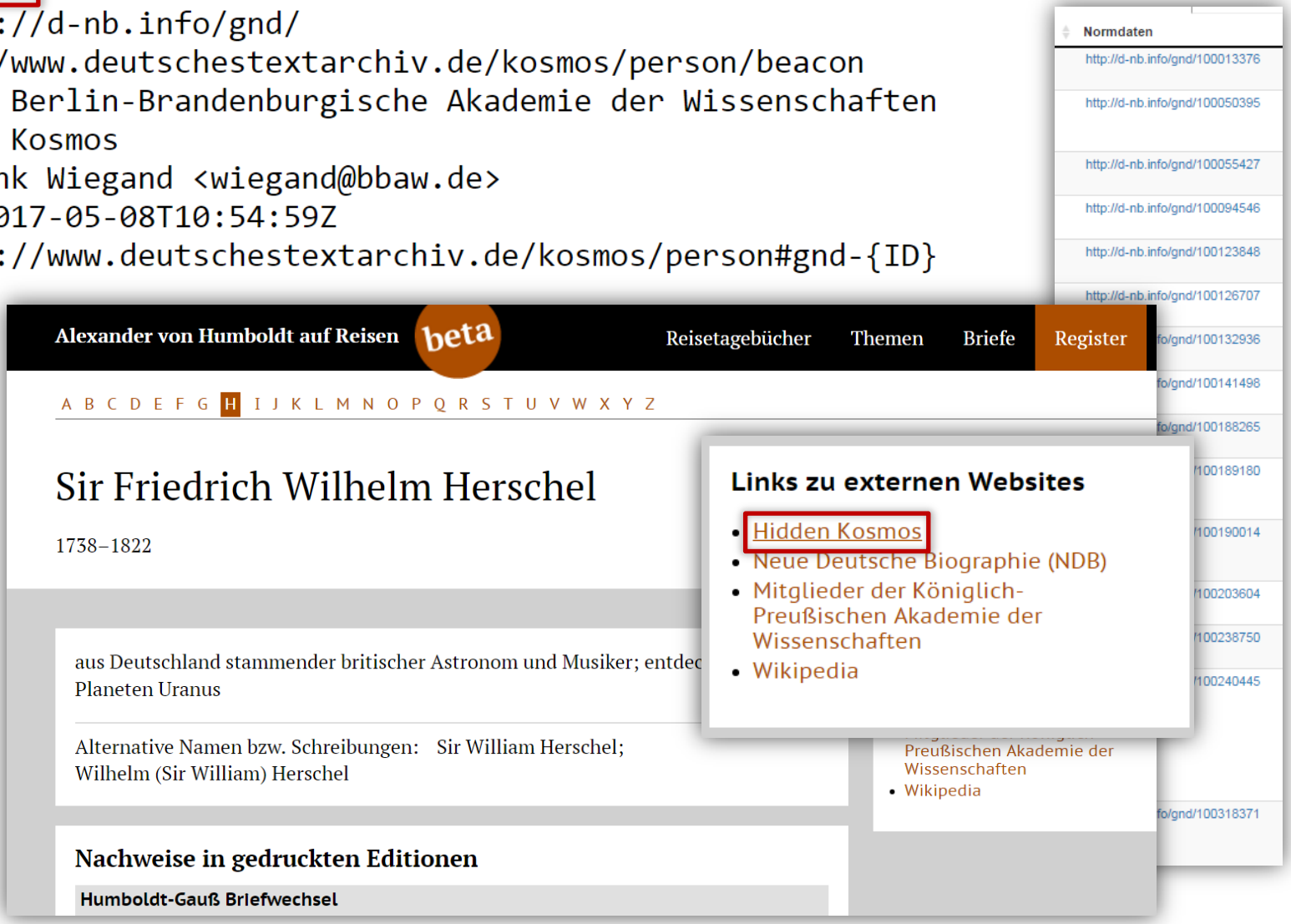
nn_oktavgfeo79_1828
Auszeichnungen mit <persName> insgesamt: 1313
<persName>s mit @ref: 1249
<persName>s ohne @ref: 64
unique Personen: 456 (+ 64 ohne @ref)

parthey_msgermqu1711_1828
Auszeichnungen mit <persName> insgesamt: 1921
<persName>s mit @ref: 1888
<persName>s ohne @ref: 33
unique Personen: 655 (+ 33 ohne @ref)

riess_f2e1853_1828
Auszeichnungen mit <persName> insgesamt: 634
<persName>s mit @ref: 622
<persName>s ohne @ref: 12
unique Personen: 257 (+ 12 ohne @ref)

Beispiel: Personen Kosmos-Vorträge

```
#FORMAT: BEACON
#PREFIX: http://d-nb.info/gnd/
#FEED: http://www.deutschestextarchiv.de/kosmos/person/beacon
#INSTITUTION: Berlin-Brandenburgische Akademie der Wissenschaften
#NAME: Hidden Kosmos
#CONTACT: Frank Wiegand <wiegand@bbaw.de>
#TIMESTAMP: 2017-05-08T10:54:59Z
#TARGET: http://www.deutschestextarchiv.de/kosmos/person#gnd-{ID}
100013376|2
100050395|5
100055427|14
100094546|1
100123848|7
100126707|4
100132936|2
100141498|2
100188265|1
100189180|7
100190014|29
100203604|2
100238750|7
100240445|23
100318371|23
100347371|4
100351212|29
```



Beispiel: Novellenschatz / Einleitung

DTAQ zuletzt gelesen · Hilfe · Zufallsseite ChristianThomas | Admin | Profil | ausloggen

sacher_kolomea_1910 (Novellenschatz)

offene Tickets: 0 (0 *ganzes Buch*)
Stand: 2017-05-05 16:28:31

Text	Text/Bild
0 – 0 – 88	0 – 0 – 88

Bild: 0008 << vorherige Seite nächste Seite >

```
1 <TEI xmlns="http://www.tei-c.org/ns/1.0"><text><front><pb
  facs="#f0008"/>
2 <div type="preface">    <p><persName ref="http://d-
  nb.info/gnd/118604570">Leopold Ritter von Sacher-
  Masoch</persName> wurde am 27. Januar 1836 in Lemberg in
  Galizien geboren, wo     sein Vater k. k. Hofrath und
  Polizeichef war. Eine kleinrussische Amme aus dem
  Bauernstande     weckte zuerst die Phantasie des Knaben
  durch die Volkslieder, die sie ihm sang, die Geschichten
  und Märchen, die sie ihm erzählte. Zunächst machten die
  blutigen Scenen der Revolution von 1846,     die Erhebung
  der Polen gegen Oesterreich und die Niedermetzelung
  derselben durch die galizischen     Bauern einen großen
  Eindruck auf ihn. Er besuchte das Gymnasium in Lemberg und
  Prag, machte     seine Universitätsstudien in Prag und
  Gratz, wurde in Gratz zum Doctor promovirt und war an der
    dortigen Hochschule durch fast zehn Jahre als Privatdocent
  der Geschichte thätig. Seine Erfolge     als Schriftsteller
  bestimmten ihn 1870 dem Lehramt den Rücken zu kehren. Im
  Jahre 1872 vermählte     er sich mit der Schriftstellerin
  <persName ref="http://d-nb.info/gnd/119392003">Wanda von
  Dunajew</persName> (<persName ref="http://d-
  nb.info/gnd/119392003">Angelika von Rümelin</persName>) und
  lebt seitdem in     Bruck an der Mur ausschließlich seinen
  literarischen Arbeiten.</p><lb/>
3 <p>In früheren Jahren bereis'te er den größten Theil
  Europa's, insbesondere lebte er mehrere     Jahre in Italien
  und Frankreich.</p><lb/>
4 <p>Seine literarische Laufbahn begann <persName
  ref="http://d-nb.info/gnd/118604570">Sacher-
  Masoch</persName> mit dem Roman:     „<bibl><ref
  target="https://books.google.de/books?id=SqMqAAAAYAAJ">Eine
  galizische Geschichte 1846</ref></bibl>“<note
  type="editorial">D. i.  <bibl>[Anonym]: Eine galizische
  Geschichte. Schaffhausen: Hurter, 1858.</bibl> Online
  verfügbar: Google Books, <ref
  target="https://books.google.de/books?
  id=SqMqAAAAYAAJ">id=SqMqAAAAYAAJ</ref>.<lb/>
5 </note> im Jahre 1858. Es folgte eine Reihe minder
  bedeutender     Productionen, unter denen das historische
  Lustspiel „<bibl>Der Mann ohne Vorurtheil</bibl>“<note
  type="editorial">Leopold Ritter von Sacher-Masoch: Der Mann
  ohne Vorurtheil. Historisches Lustspiel in fünf Acten.
  Leipzig: Leiner (= Deutsche Schaubühne), 1866.</note> den
  meisten     Erfolg hatte; in Deutschland lenkte zuerst<lb/>
6 </p></div></front></text></TEI>
```

Leopold Ritter von Sacher-Masoch wurde am 27. Januar 1836 in Lemberg in Galizien geboren, wo sein Vater k. k. Hofrath und Polizeichef war. Eine kleinrussische Amme aus dem Bauernstande weckte zuerst die Phantasie des Knaben durch die Volkslieder, die sie ihm sang, die Geschichten und Märchen, die sie ihm erzählte. Zunächst machten die blutigen Scenen der Revolution von 1846, die Erhebung der Polen gegen Oesterreich und die Niedermetzelung derselben durch die galizischen Bauern einen großen Eindruck auf ihn. Er besuchte das Gymnasium in Lemberg und Prag, machte seine Universitätsstudien in Prag und Gratz, wurde in Gratz zum Doctor promovirt und war an der dortigen Hochschule durch fast zehn Jahre als Privatdocent der Geschichte thätig. Seine Erfolge als Schriftsteller bestimmten ihn 1870 dem Lehramt den Rücken zu kehren. Im Jahre 1872 vermählte er sich mit der Schriftstellerin Wanda von Dunajew (Angelika von Rümelin) und lebt seitdem in Bruck an der Mur ausschließlich seinen literarischen Arbeiten.

In früheren Jahren bereis'te er den größten Theil Europa's, insbesondere lebte er mehrere Jahre in Italien und Frankreich.

Seine literarische Laufbahn begann Sacher-Masoch mit dem Roman: „Eine galizische Geschichte 1846“ im Jahre 1858. Es folgte eine Reihe minder bedeutender Productionen, unter denen das historische Lustspiel „Der Mann ohne Vorurtheil“ den meisten Erfolg hatte; in Deutschland lenkte zuerst

Buchdaten
DTA-Informationen
Metadaten
Ansichten (Webversion)

Korrekturstatus
Text von mir kontrolliert
Text/Bild von mir kontrolliert
Darstellung von mir kontrolliert
TEI-XML von mir kontrolliert

Tickets für diese Seite
neu: Ticket

Suche
DDC
grep

Anmerkung zu dieser Seite anlegen

→ Weitere Aufgabe: Bibliographie erwähnter Werke im Novellenschatz?

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Korpusanalyse im DTA

Suche im Deutschen Textarchiv

Treffer 6 - 15 von 41

KWIC · Neue Suche · Vorherige · Nächste · Hilfe ehelichen with

10 Treffer pro Seite Sortierung: Datum aufsteigend/abstei

6: Zesen, Philipp von: Assenat. Amsterd

Ob nun der Potifar / der den Josef ge
Potifar gewesen / dessen Tochter er
Priester zu On / und Josef der Geschi
Priester zu Heliopel / welches einerle
unterschiedliche meinungen.

7: Eckarth, Gotthülff Traugott: Chronica
Beschreibung des Dorffes Bertzdorff,
Zittau in der Ober-Lausitz gelegen. H

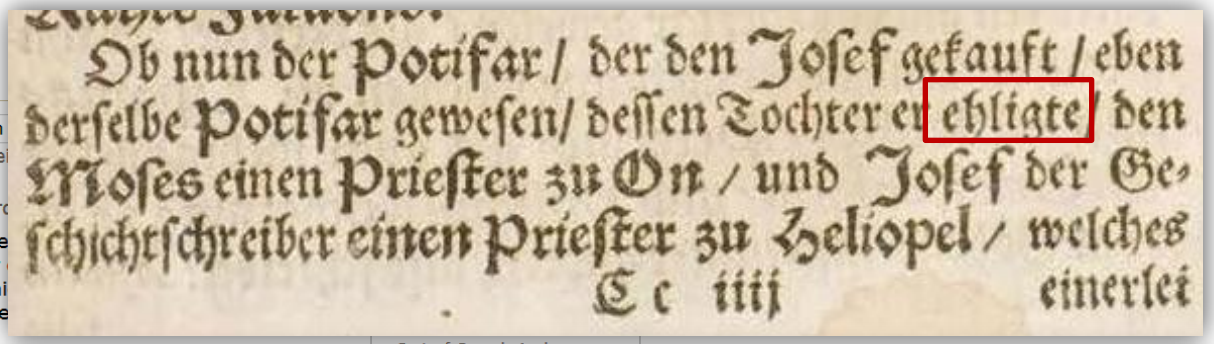
Johanna Elisabeth Rückerin, Tit. Hrn.
des Gymnasii und vornehmen Bürger
eintzigen Tochter, welche Ehe GOtt n
hat, wovon noch 2. Söhne und 3. Töc

8: Eckarth, Friedrich: Chronica Oder Historische Beschreibung [mehr]
Des Dorffes Herwigsdorff. 1737. #46

Er **ehlichte** An. 1591. den 3 Oct. zu
Hrn. Gregorii Colbenschlägers Tochte
Heimführung allhier den 28. Oct. GO
An. 1592. den 9 Sept. mit einem Soh
und den 4 Febr.

9: Albertinus, Aegidius: Der Landtstörtz
oder Picaro genannt. Bd. 1. München

Ja was mehr ist / er fragte nichts nac
sondern er **ehelichte** sie auch offent
vrsach / dann er wuste ein ansehenli
das Gelt war gewunnen / also ist es
nach dem mein Vatter Hochzeit mit
fing er an / sich sehr stattlich zu halten / zu bancketiren / zu
spielen / vnd dermassen zu dommiren / daß er jnner halb
zwey Jaren nit allein 10000 Ducaten meines Mütterlichen Guts
hindurch jagte / sonder auch sein Landtgut Alfarche genannt /
verkauffte / vnnd bald hernacher auß Hertzleyd starb.



p: Part-of-Speech-Analyse

Ob nun der Potifar / der den Josef gekauft / eben
derselbe Potifar gewesen / dessen Tochter er ehelichte / den
Moses einen Priester zu On / und Josef der
Geschichtsschreiber einen Priester zu Heliobel / welches

- Tokenisierung,
  Lemmatisierung,
  POS-Tagging,
  orthographische
  Normierung

alt:	ehligte
neu:	ehelichte

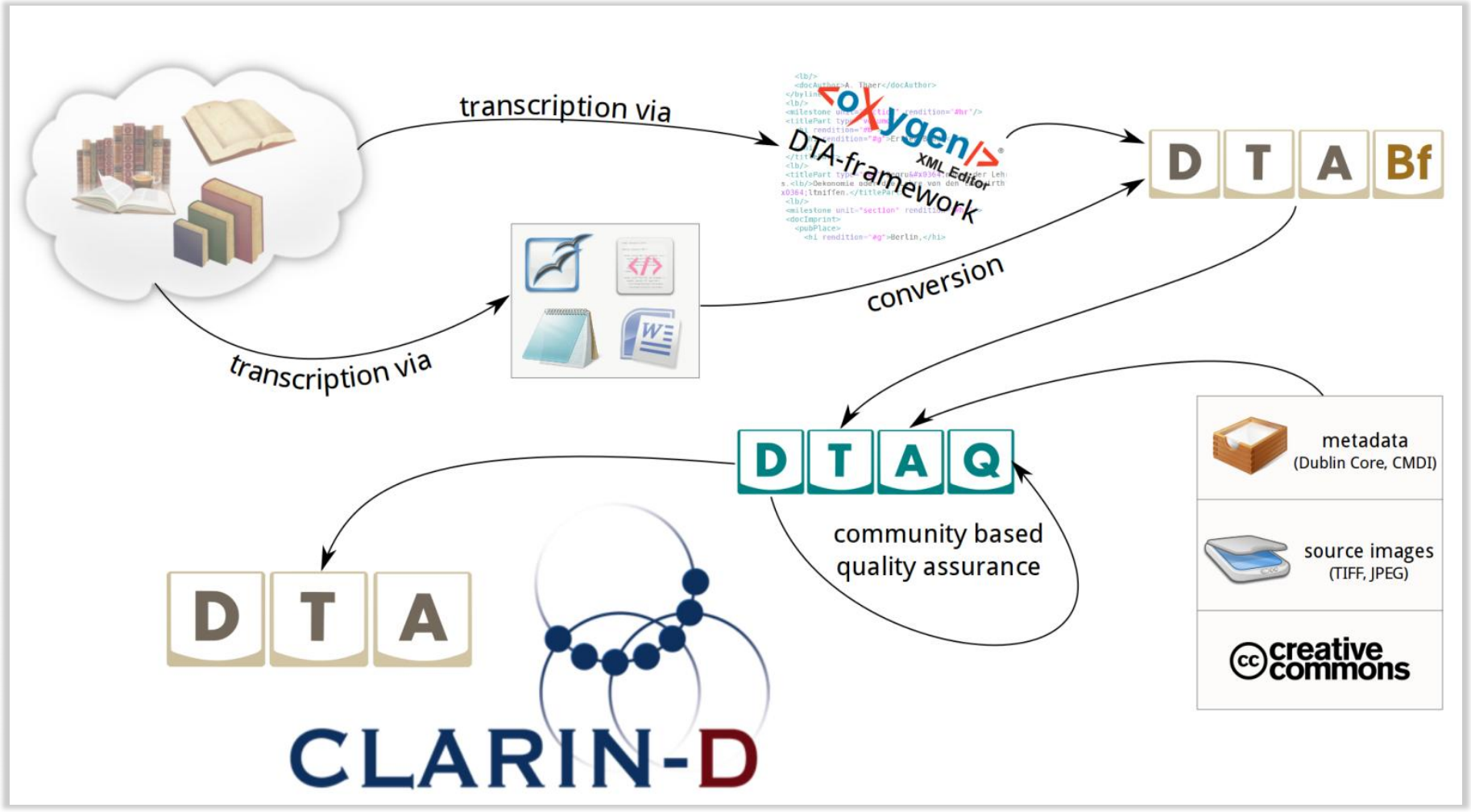
POS-Informationen

POS-Tag:	VVFIN
Lemma:	ehelichen

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Workflows im DTA



Datenmanagement im DTA

- Metadaten:
  - Metadatenbanken → zentrale DB (PostgreSQL)
  - ausführliche Informationen zu den Werken
    (Autor, Titel, Ort/Jahr, Signatur, Bearbeiter, Textklasse, ...)
  - Stationen im Workflow/Bearbeiter
  - Output in verschiedenen Formaten (TEI-Header, CMDI, DC)

Datenmanagement im DTA

- Metadaten:
  - Metadatenbanken → zentrale DB (PostgreSQL)
  - ausführliche Informationen zu den Werken (Autor, Titel, Ort/Jahr, Signatur, Bearbeiter, Textklasse, ...)
  - Stationen im Workflow/Bearbeiter
  - Output in verschiedenen Formaten (TEI-Header, CMDI, DC)
- Textdaten:
  - XML-Dokumente im Dateisystem
  - Format-Ansichten bzw. -Ausgaben basierend DTABf-XML (TCF, HTML, TXT, epub, PDF (TEI-publisher))
  - unicruftxml, normxml, lemmaxml (für Voyant u.a. Tools)
  - Analyseansichten/-ausgaben (z.B. CAB, TCF)

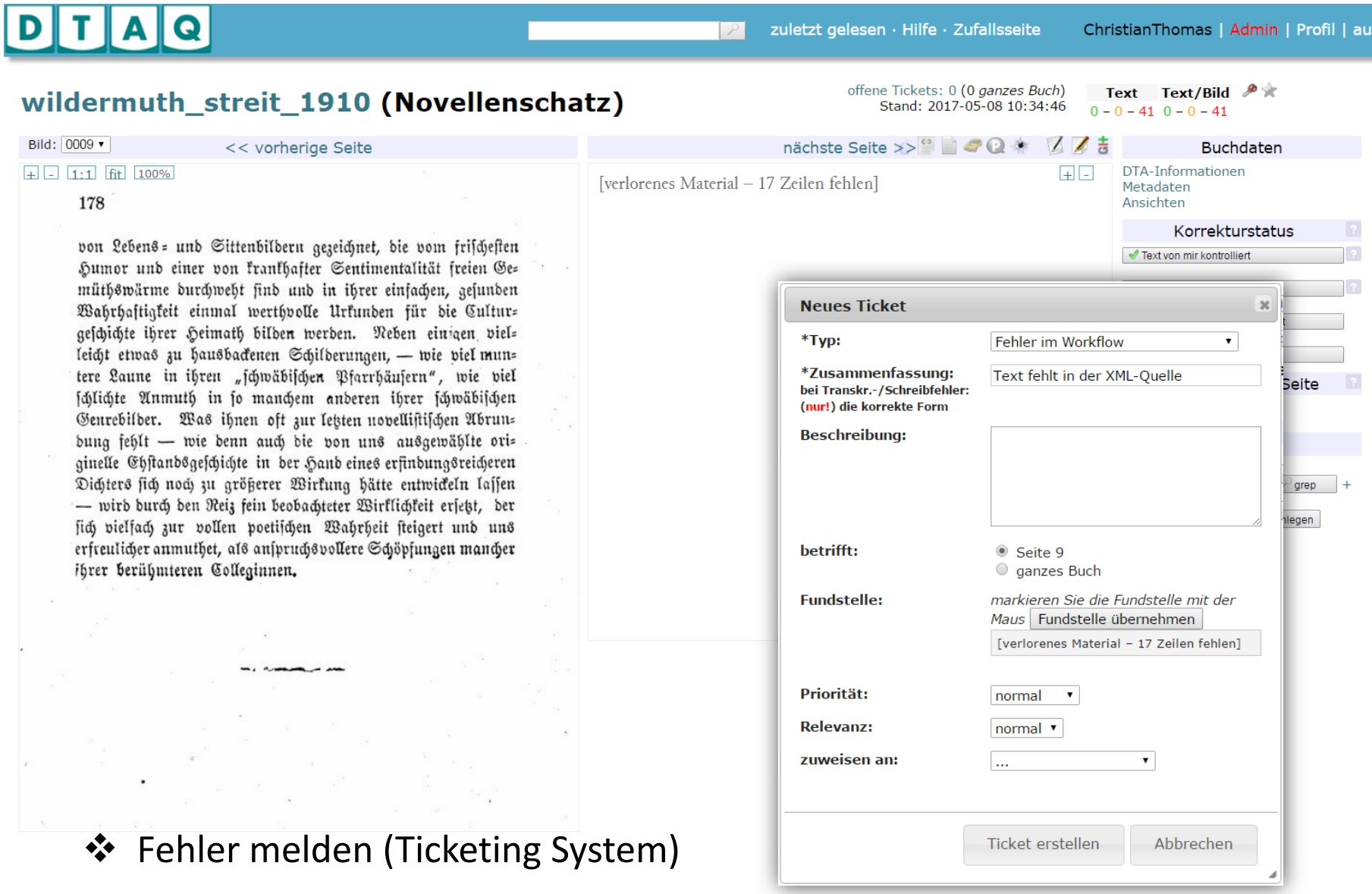
Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Datenmanagement im DTA

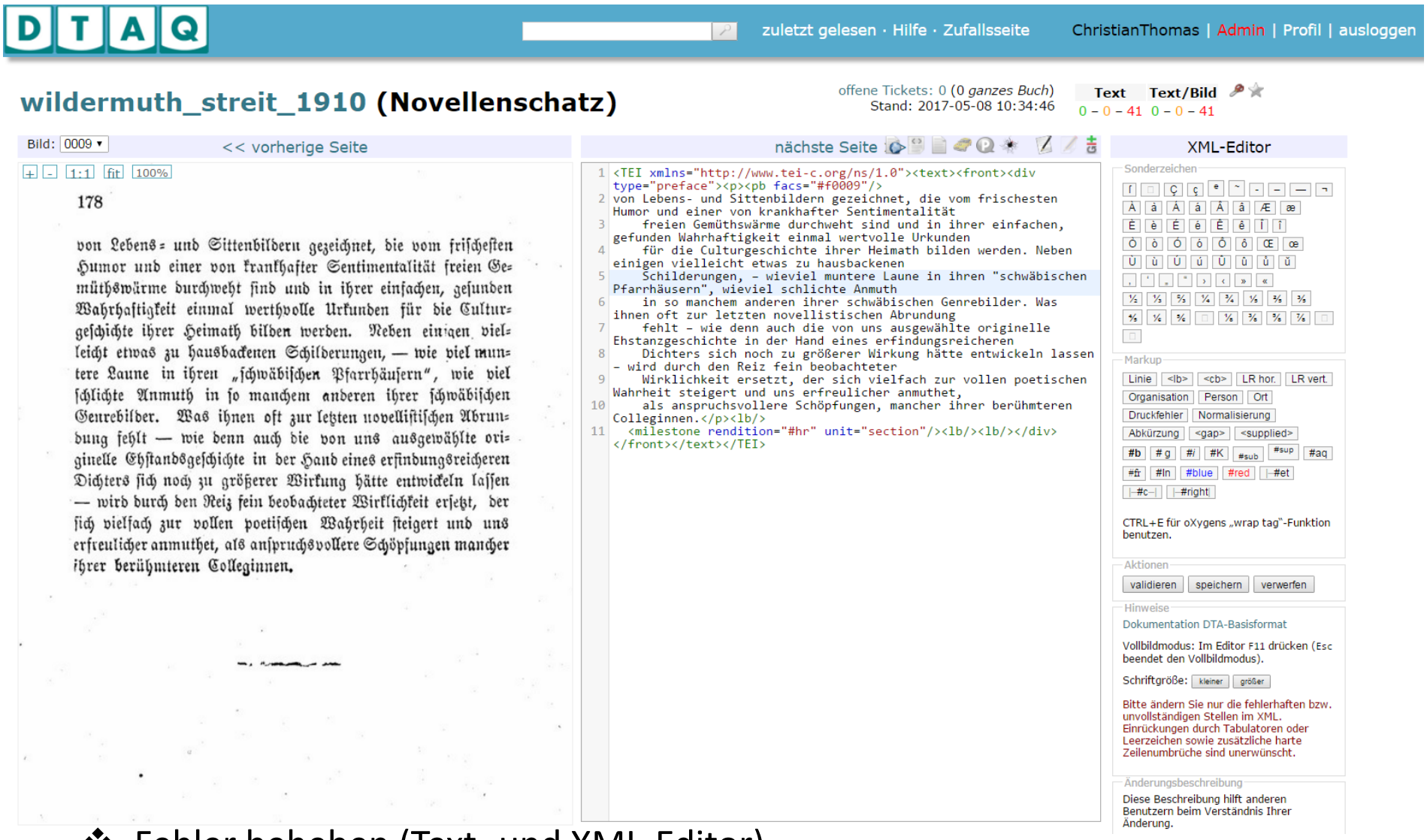
- Funktionalitäten in DTAQ
  - zentrale DB (PostgreSQL)
  - Benutzerverwaltung, Rollensystem, Rechtemanagement
  - Tickets und deren Historie
  - gelesene Seiten
  - Text- und XML-Editor
  - Tools zur Verarbeitung von DTABf-XML-Dateien u.v.m.
- Bilddaten:
  - hochaufgelöste TIFFs + Derivate im Dateisystem

Kollaboratives Arbeiten im DTA



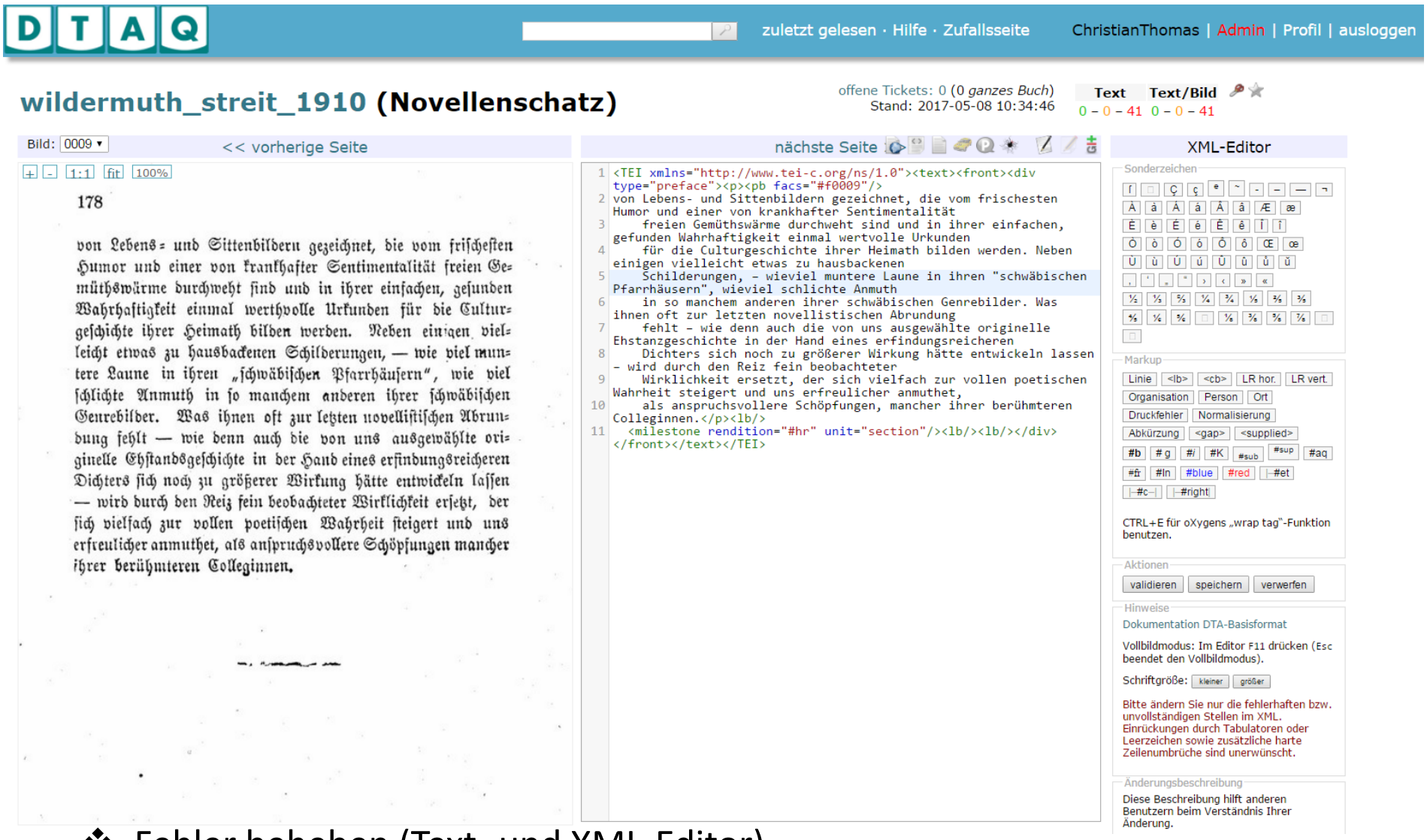
- Fehler melden (Ticketing System)

Kollaboratives Arbeiten im DTA



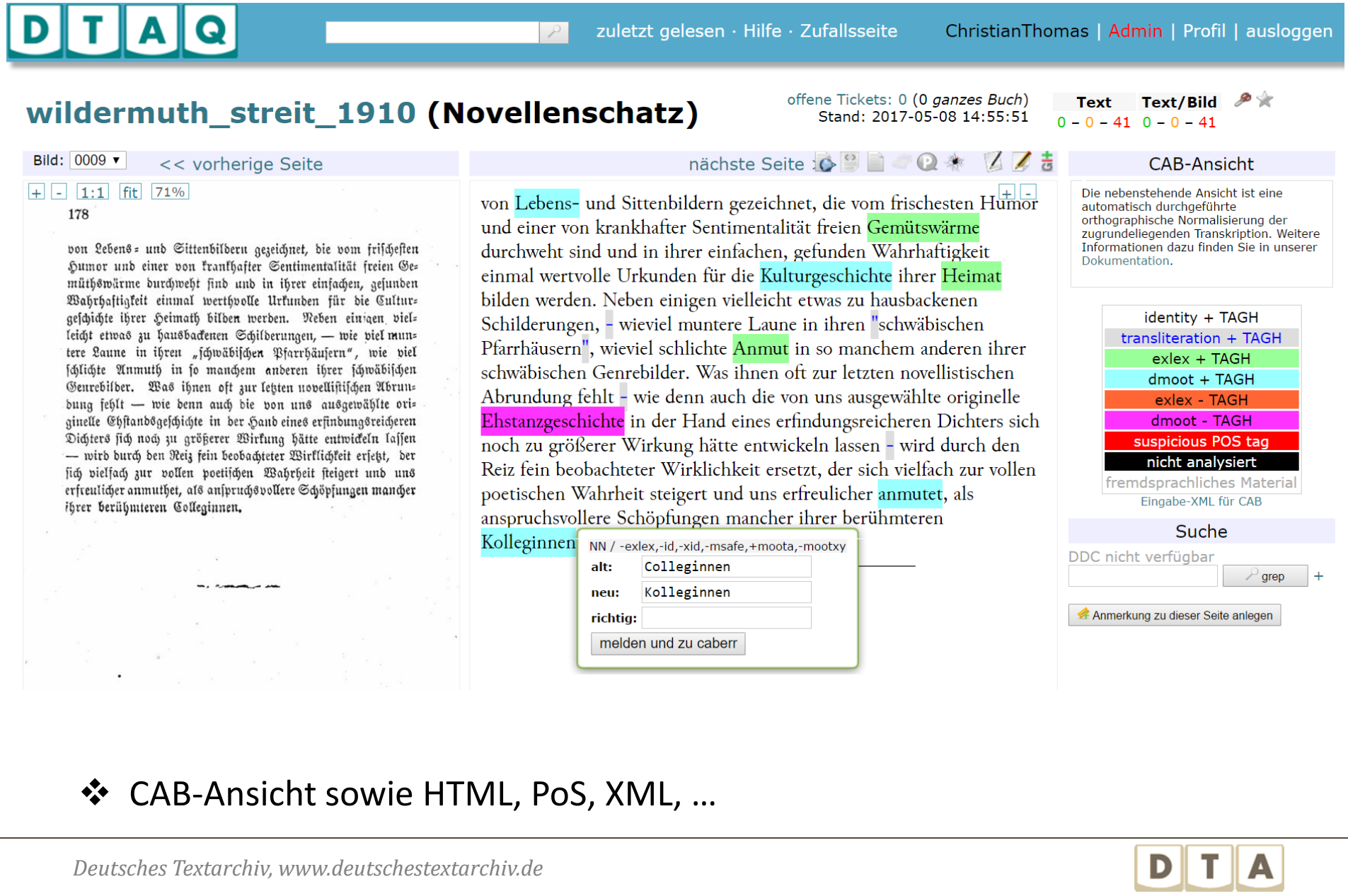
- Fehler beheben (Text- und XML-Editor)

Kollaboratives Arbeiten im DTA



- Fehler beheben (Text- und XML-Editor)

Kollaboratives Arbeiten im DTA



- CAB-Ansicht sowie HTML, PoS, XML, …

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Die BBAW als CLARIN-Zentrum

CLARIN ERIC: Common Language Resources and Technology Infrastructure

- CLARIN-D = deutscher Partner in CLARIN ERIC
- Zentrum Sprache der BBAW = CLARIN-D-Zentrum
- **DTA als Forschungsplattform für historische Daten**

Die BBAW als CLARIN-Zentrum

CLARIN ERIC: Common Language Resources and Technology Infrastructure

- CLARIN-D = deutscher Partner in CLARIN ERIC
- Zentrum Sprache der BBAW = CLARIN-D-Zentrum
- **DTA als Forschungsplattform für historische Daten**
- Nachhaltige Datenaufbewahrung im Repositorium
- Bereitstellung aller Daten in CLARIN-Formaten
  → CLARIN-Infrastruktur (VLO, FCS, WebLicht, …)
- Ressourcenkuration, Schulungen, Nutzerbetreuung, Dokumentation, Use Cases

Aufbau des Vortrags

1. Zum Deutschen Textarchiv
   - ❖ Aufbereitung der Daten im Deutschen Textarchiv
   - ❖ Möglichkeiten der Korpusanalyse im DTA
2. Workflows und Datenmanagement
   - ❖ kollaborative Arbeitsumgebung DTAQ
3. Nutzung und Vernetzung von Daten über das DTA
   - ❖ über das DTA in die CLARIN-Infrastruktur
   - ❖ Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

Das Korpus „Novellenschatz“ im DTA-Kontext

Digital Humanities Cooperation der Technischen Universität Darmstadt und der Universität Konstanz

Das Korpus „Deutscher Novellenschatz“ ist lizenziert unter der Namensnennung 4.0 International (CC BY 4.0) Lizenz.

Weiterführende Informationen

Der deutsche Novellenschatz, herausgegeben von Paul Heyse und Hermann Kurz

In Vorarbeit für das von der VolkswagenStiftung geförderte Projekt *Reading at Scale. Mixing Methods in Literary Corpus Analysis* wurde der *Deutsche Novellenschatz* (24 Bände, 1871–1876) von der *Digital Humanities Cooperation* der Technischen Universität Darmstadt und der Universität Konstanz unter Leitung von Prof. Dr. Thomas Weitin in einem corrected OCR-Verfahren digitalisiert, im XML/TEI-Format codiert und mit Metadaten angereichert. Neben 86 teils kanonischen, teils unbekannten Novellen von 82 AutorInnen enthält das Korpus kurze Einführungen zu jedem Text sowie einen als Einleitung in die Sammlung dienenden gattungspoetischen Essay zur Novellenform, der durch seine so genannte Falkentheorie bekannt geworden ist. Innerhalb des Projekts *Reading at Scale* sollen weitere realistische Novellensammlungen, z. B. der *Neue deutsche Novellenschatz* (24 Bände, 1884–1887), aufbereitet werden.

Dokumentliste

Dokument	Größe	Titel
alexis_sacken_1910	178.99 KB	Alexis, Willibald: Herr von Sacken. In: Deuts… Berlin, [1910], S. 95–202.
andolt_nacht_1910	152.01 KB	Andolt, Ernst [d. i. Bernhard Abeken]: Eine N… 10. 2. Aufl. Berlin, [1910], S. 211–287.
arnim_invalide_1910	76.57 KB	Arnim, Achim von: Der tolle Invalide auf de… Hermann Kurz. Bd. 1. Berlin, [1871], S. [163]–…
auerbach_diethelm_1910	513.92 KB	Auerbach, Berthold: Die Geschichte des Di… Hermann Kurz. Bd. 7. 2. Aufl. Berlin, [1910], S…
berthold_irrwischfritze_1910	190.53 KB	Berthold, Franz [d. i. Adelheid Reinbold]: Irrwisch… …ann Kurz. Bd. 4. 2. Aufl. Berlin, [1910], S. [1]–115.
brentano_annerl_1910	92.54 KB	Brentano, Clemens: Geschichte vom braven Kasperl und dem schönen Annerl. In: Deutscher Novellenschatz. Hrsg. von Paul Heyse und Hermann Kurz. Bd. 1. München, [1871], S. [107]–162.

- 87 Dokumente
- ca. 7.700 Seiten
- 1.936.327 Tokens
- 10.809.093 Zeichen

Das Korpus „Novellenschatz“ im DTA-Kontext

Novellenschatz XML

```
<ab>Wo ist die Hand so zart, daß ohne Irren</ab>
<ab>Sie sondern mag beschränkten Hirnes Wirren,</ab>
<ab>So fest, daß ohne Zittern sie den Stein</ab>
<ab>Mag schleudern auf ein arg verkümmert Sein?</ab>
<ab>Wer wagt es, eitlen Blutes Drang zu messen,</ab>
<ab>Zu wägen jedes Wort, das unvergessen</ab>
<ab>In junge Brust die zähen Wurzeln trieb,</ab>
<ab>Des Vorurtheils geheimen Seelendieb?</ab>
<ab>Du Glücklicher, geboren und gehegt</ab>
<ab>Im lichten Raum, von frommer Hand gepflegt.</ab>
<ab>Leg hin die Wagschal', nimmer dir erlaubt!</ab>
<ab>Laß ruhn den Stein — er trifft dein eignes</ab>
<ab>Haupt! —</ab>
```

Novellenschatz Vorlage

Wo iſt die Hand ſo zart, daß ohne Irren
Sie ſondern mag beſchränkten Hirnes Wirren,
So feſt, daß ohne Zittern ſie den Stein
Mag ſchleudern auf ein arm verkümmert Sein?
Wer wagt es, eitlen Blutes Drang zu meſſen,
Zu wägen jedes Wort, das unvergeſſen
In junge Bruſt die zähen Wurzeln trieb,
Des Vorurtheils geheimen Seelendieb?
Du Glücklicher, geboren und gehegt
Im lichten Raum, von frommer Hand gepflegt,
Leg hin die Wagſchal', nimmer dir erlaubt!
Laß ruhn den Stein — er trifft dein eignes
Haupt! —

Friederich Mergel, geboren 1738, war der Sohn eines ſogenannten Halbmeiers oder Grundeigenthümers geringer Klaſſe im Dorfe B., das, ſo ſchlecht gebaut und

droste_judenbuche_1910

Das Korpus „Novellenschatz“ im DTA-Kontext

Novellenschatz XML

```
<ab>Wo ist die Hand so zart, daß ohne Irren</ab>
<ab>Sie sondern mag beschränkten Hirnes Wirren,</ab>
<ab>So fest, daß ohne Zittern sie den Stein</ab>
<ab>Mag schleudern auf ein arg verkümmert Sein?</ab>
<ab>Wer wagt es, eitlen Blutes Drang zu messen,</ab>
<ab>Zu wägen jedes Wort, das unvergessen</ab>
<ab>In junge Brust die zähen Wurzeln trieb,</ab>
<ab>Des Vorurtheils geheimen Seelendieb?</ab>
<ab>Du Glücklicher, geboren und gehegt</ab>
<ab>Im lichten Raum, von frommer Hand gepflegt.</ab>
<ab>Leg hin die Wagschal', nimmer dir erlaubt!</ab>
<ab>Laß ruhn den Stein — er trifft dein eignes</ab>
<ab>Haupt! —</ab>
```

DTABf XML

```
<epigraph>
  <lg type="poem">
    <l> Wo ist die Hand so zart, daß ohne Irren</l>
    <l> Sie sondern mag beschränkten Hirnes Wirren,
    <l> So fest, daß ohne Zittern sie den Stein</l>
    <l> Mag schleudern auf ein arg verkümmert Sein
    <l>Wer wagt es, eitlen Blutes Drang zu messen,
    <l>Zu wägen jedes Wort, das unvergessen</l><lb
    <l>In junge Brust die zähen Wurzeln trieb,</l>
    <l>Des Vorurtheils geheimen Seelendieb?</l><lb
    <l>Du Glücklicher, geboren und gehegt</l><lb/>
    <l>Im lichten Raum, von frommer Hand gepflegt.
    <l>Leg hin die Wagschal', nimmer dir erlaubt!</
    <l>Laß ruhn den Stein — er trifft dein eignes
    <l>Haupt! —</l><lb/>
  </lg>
</epigraph>
```

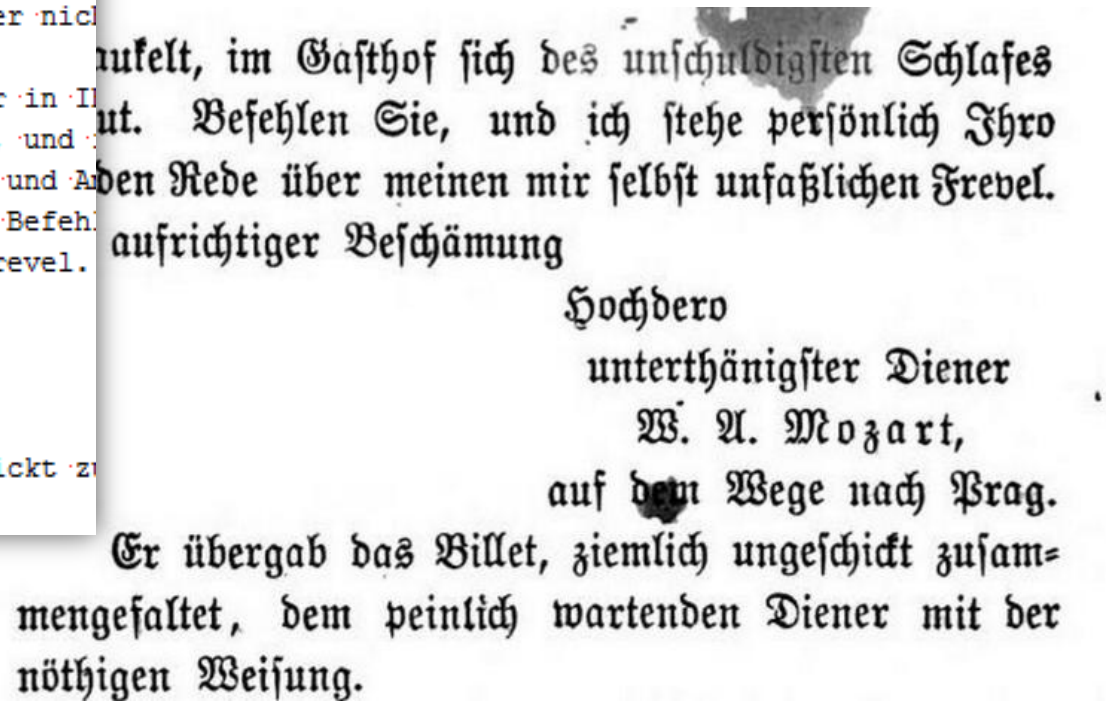
droste_judenbuche_1910

Das Korpus „Novellenschatz“ im DTA-Kontext

Novellenschatz XML

```
<p>Zwei Gartenknechte kamen nun wirklich herbei,
hinweg. Inzwischen hatte unser Meister seine Br
herausgenommen, und während daß der Gärtner nicl
zu schreiben:</p>
<p>Gnädigste Frau! Hier sitze ich Unseliger in I
Apfel gekostet. Das Unglück ist geschehen, und
Eva schieben, die eben jetzt, von Grazien und A
sich des unschuldigsten Schlafes erfreut. Befeh
Rede über meinen mir selbst unfaßlichen Frevel.
<ab>Hochdero</ab>
<ab>unterthänigster Diener</ab>
<ab>W. A. Mozart,</ab>
<ab>auf dem Wege nach Prag.</ab>
<p>Er übergab das Billet, ziemlich ungeschickt z
mit der nöthigen Weisung.</p>
```

Novellenschatz Vorlage



ukelt, im Gasthof sich des unschuldigsten Schlafes
ut. Befehlen Sie, und ich stehe persönlich Ihro
den Rede über meinen mir selbst unfaßlichen Frevel.
aufrichtiger Beschämung

Hochdero
unterthänigster Diener
W. A. Mozart,
auf dem Wege nach Prag.

Er übergab das Billet, ziemlich ungeschickt zusam-
mengefaltet, dem peinlich wartenden Diener mit der
nöthigen Weisung.

moerike_mozart_1910

Das Korpus „Novellenschatz“ im DTA-Kontext

Novellenschatz XML

```
<p>Zwei Gartenknechte kamen nun wirklich herbei,
hinweg. Inzwischen hatte unser Meister seine Br
herausgenommen, und während daß der Gärtner nicl
zu schreiben:</p>
<p>Gnädigste Frau! Hier sitze ich Unseliger in I
Apfel gekostet. Das Unglück ist geschehen, und
Eva schieben, die eben jetzt, von Grazien und A
sich des unschuldigsten Schlafes erfreut. Befeh
Rede über meinen mir selbst unfaßlichen Frevel.
<ab>Hochdero</ab>
<ab>unterthänigster Diener</ab>
<ab>W. A. Mozart,</ab>
<ab>auf dem Wege nach Prag.</ab>
<p>Er übergab das Billet, ziemlich ungeschickt z
mit der nöthigen Weisung.</p>
```

DTABf XML

```
<p>Zwei Gartenknechte kamen nun wirklich herbei, luc
hinweg. Inzwischen hatte unser Meister seine Brieft
herausgenommen, und während daß der Gärtner nicht
zu schreiben:</p><lb/>
<div type="letter">
<p>Gnädigste Frau! Hier sitze ich Unseliger in Ihren
Apfel gekostet. Das Unglück ist geschehen, und ich
Eva schieben, die eben jetzt, von Grazien und Amore
sich des unschuldigsten Schlafes erfreut. Befehlen
Rede über meinen mir selbst unfaßlichen Frevel. Mit
<closer>
<salute>Hochdero
unterthänigster Diener</salute>
<signed>W. A. Mozart,</signed>
auf dem Wege nach Prag.</closer>
</div>
<div>
<p>Er übergab das Billet, ziemlich ungeschickt zusan
mit der nöthigen Weisung.</p><lb/>
```

moerike_mozart_1910

Das Korpus „Novellenschatz“ im DTA-Kontext

Novellenschatz XML

```
    <p>Es bleibt nichts übrig, a
        sind, wegzudenken: dann t
        Tage, die einen großartig
    <ab>K.</ab>
</front>
```

```
        versiegt war. Zwei Jahre
        einzelne Geschichten in
        Zeit der Name des Dichte
    <ab>H.</ab>
</front>
```

DTABf XML

```
    <p>Es bleibt nichts übrig, als diese kleinen Auswüchse, die ja nicht
        sind, wegzudenken: dann tritt in den wenigen Blättern eine höchst
        Tage, die einen großartigen und rührenden Eindruck hinterlassen
        <byline>
            <persName ref="http://d-nb.info/gnd/118778277">K.</persName>
        </byline><lb/>
        <pb facs="#f0011"/>
    </div>
</front>
```

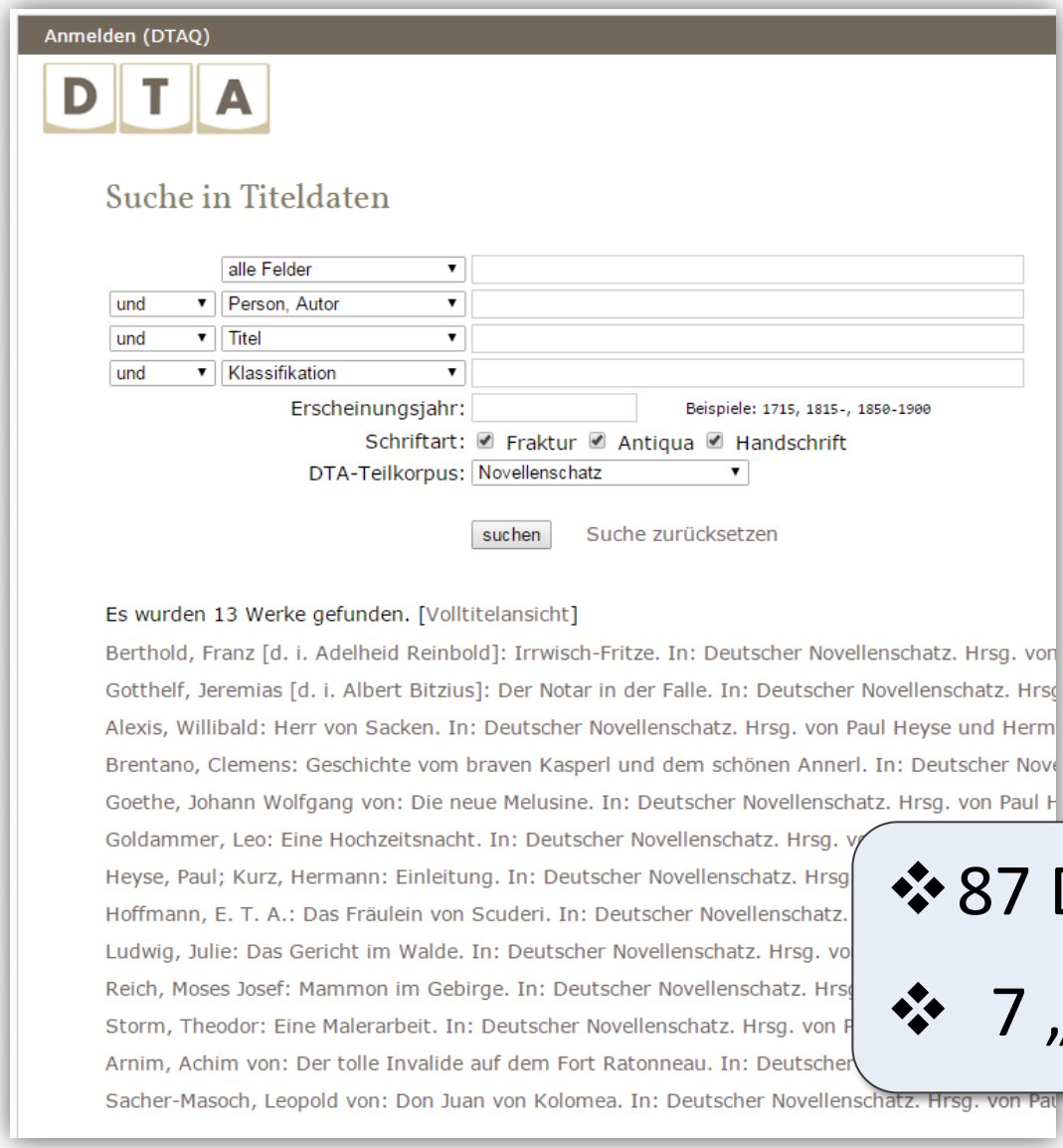
arnim_invalide_1910

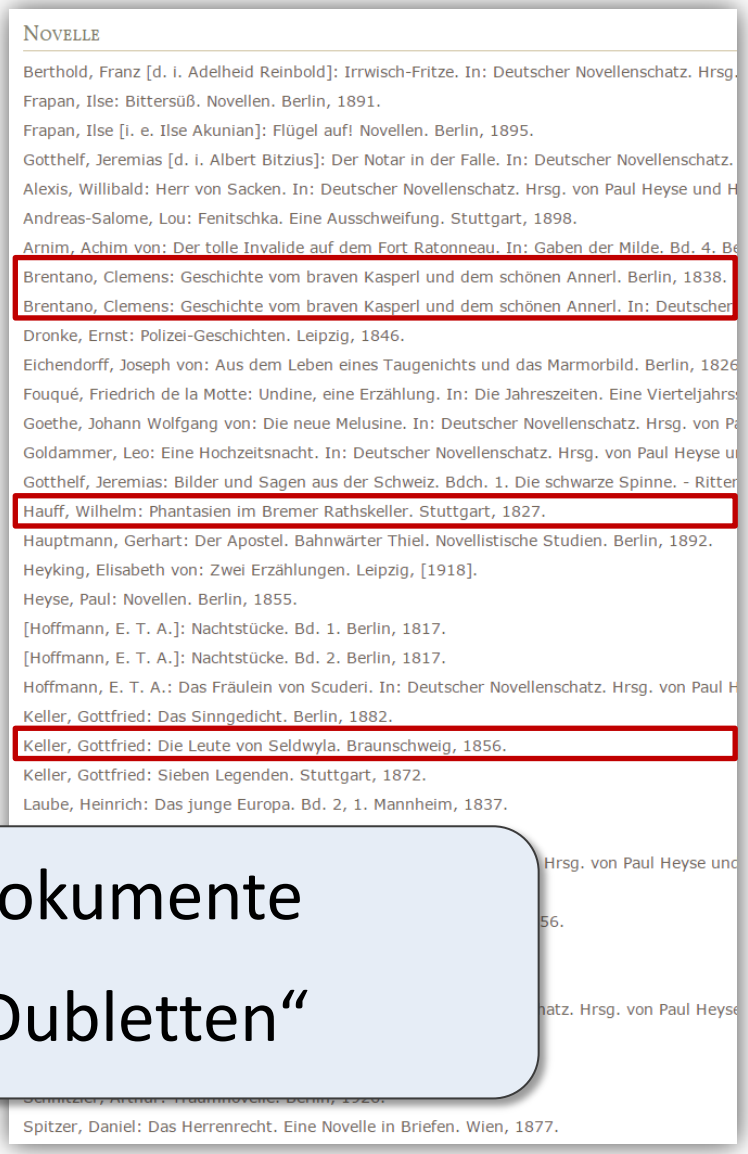
```
        einzelne Geschichten in Auerbach's Volkskalender, so daß in un
        Zeit der Name des Dichters noch lange nicht nach Verdienst gek
    <byline>
        <persName ref="http://d-nb.info/gnd/118550772">H.</persName>
    </byline><lb/>
    </div>
</front>
```

keller_dorfe_1910

→ Weitere Aufgabe: Identifikation der Verfasserschaft nicht gekennzeichneter Einleitungen?

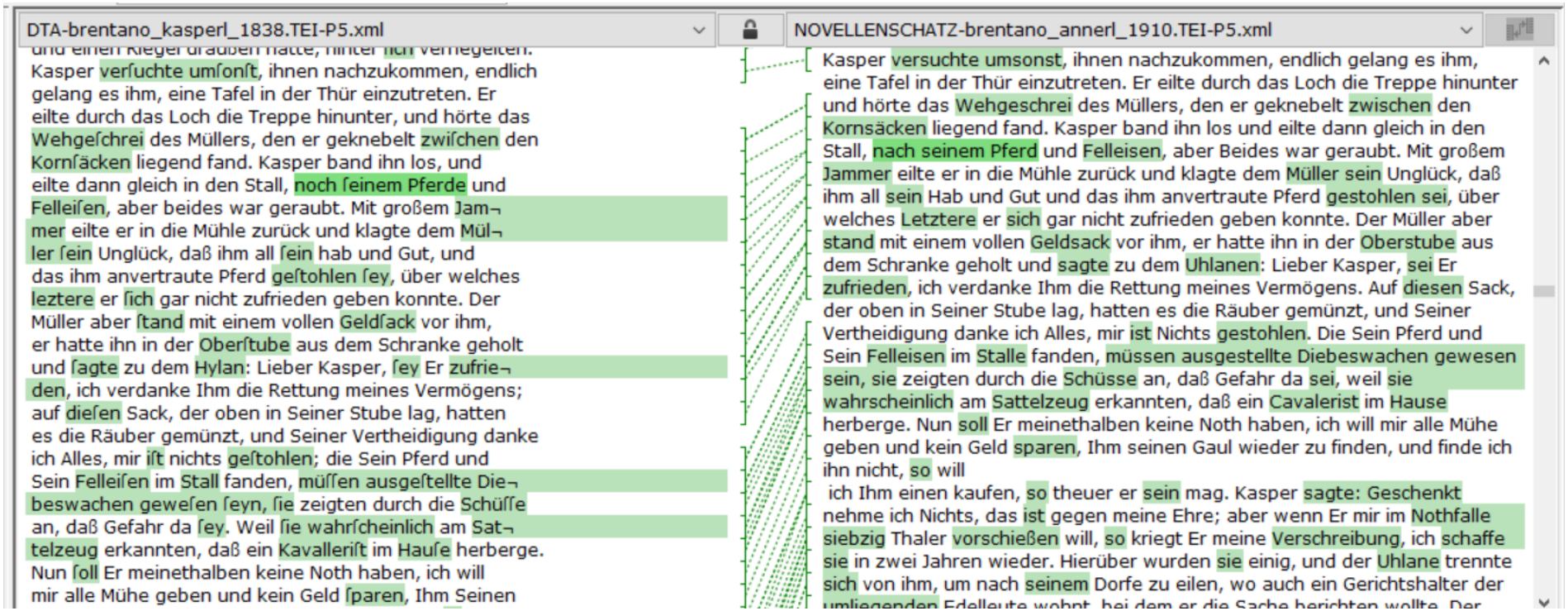
Das Korpus „Novellenschatz“ im DTA-Kontext



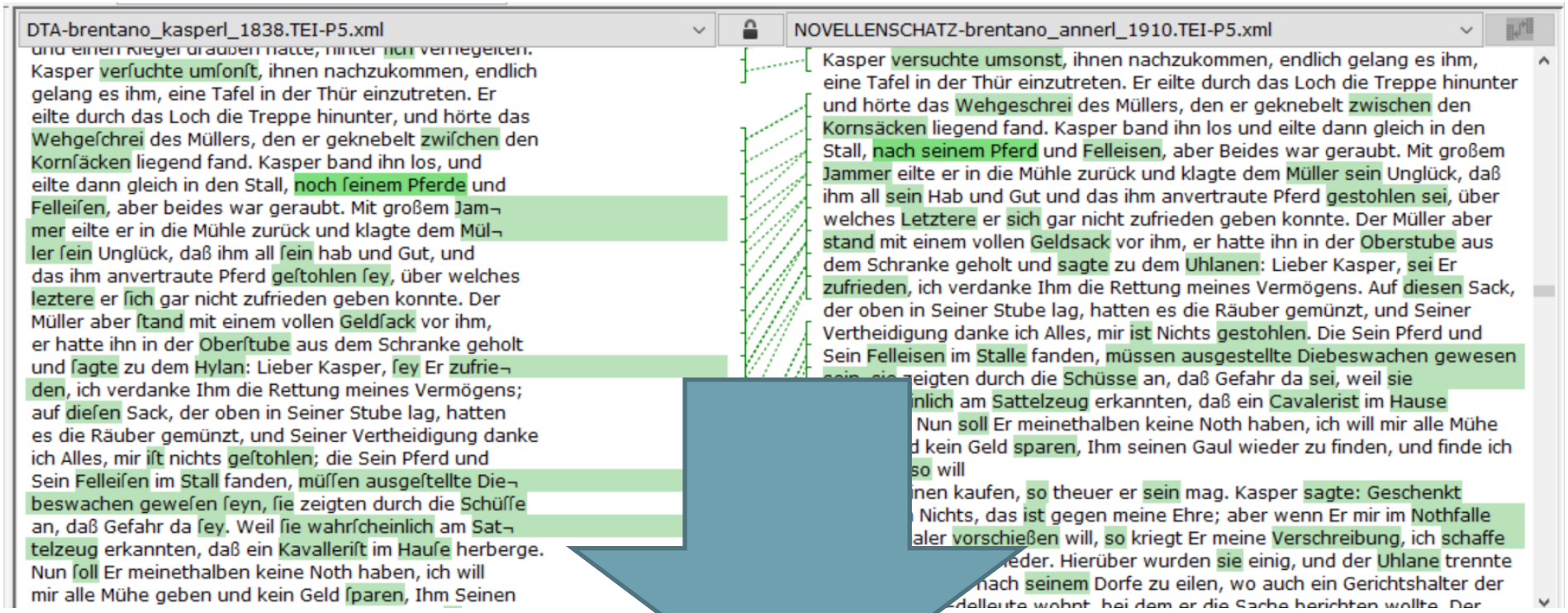


- 87 Dokumente
- 7 „Dubletten“

Kollation zweier Ausgaben mit juXta (Brentano)



Kollation zweier Ausgaben mit juXta (Brentano)



Texte normalisieren

Dieses Tool konvertiert eine DTABf-konforme Eingabedatei dahingehend, dass alle Wortformen auf moderne Schreibweisen normalisiert werden.

Datei auswählen Keine ausgewählt

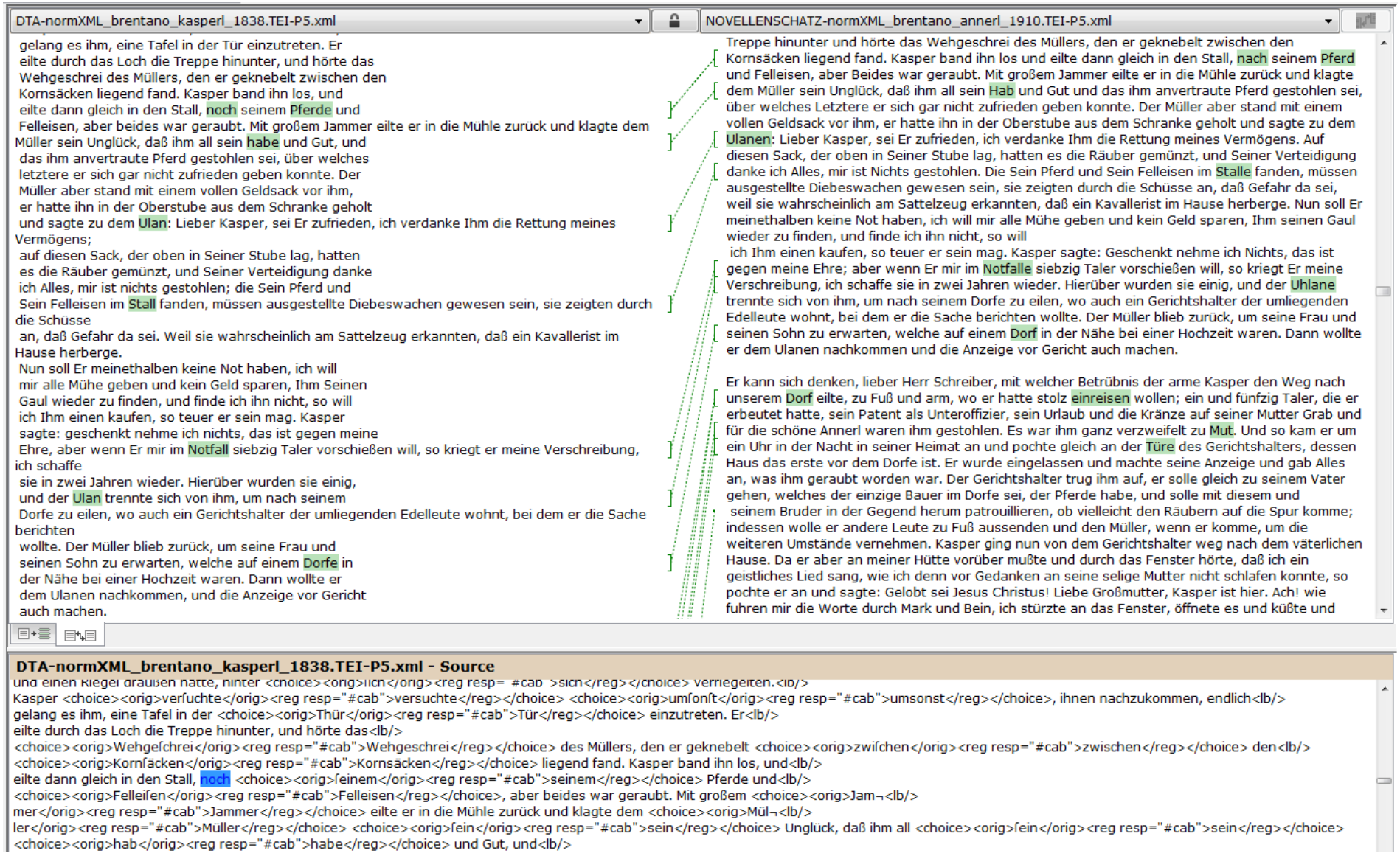
Ergebnis ◉ anzeigen ○ zum Download Datei verarbeiten

Skript

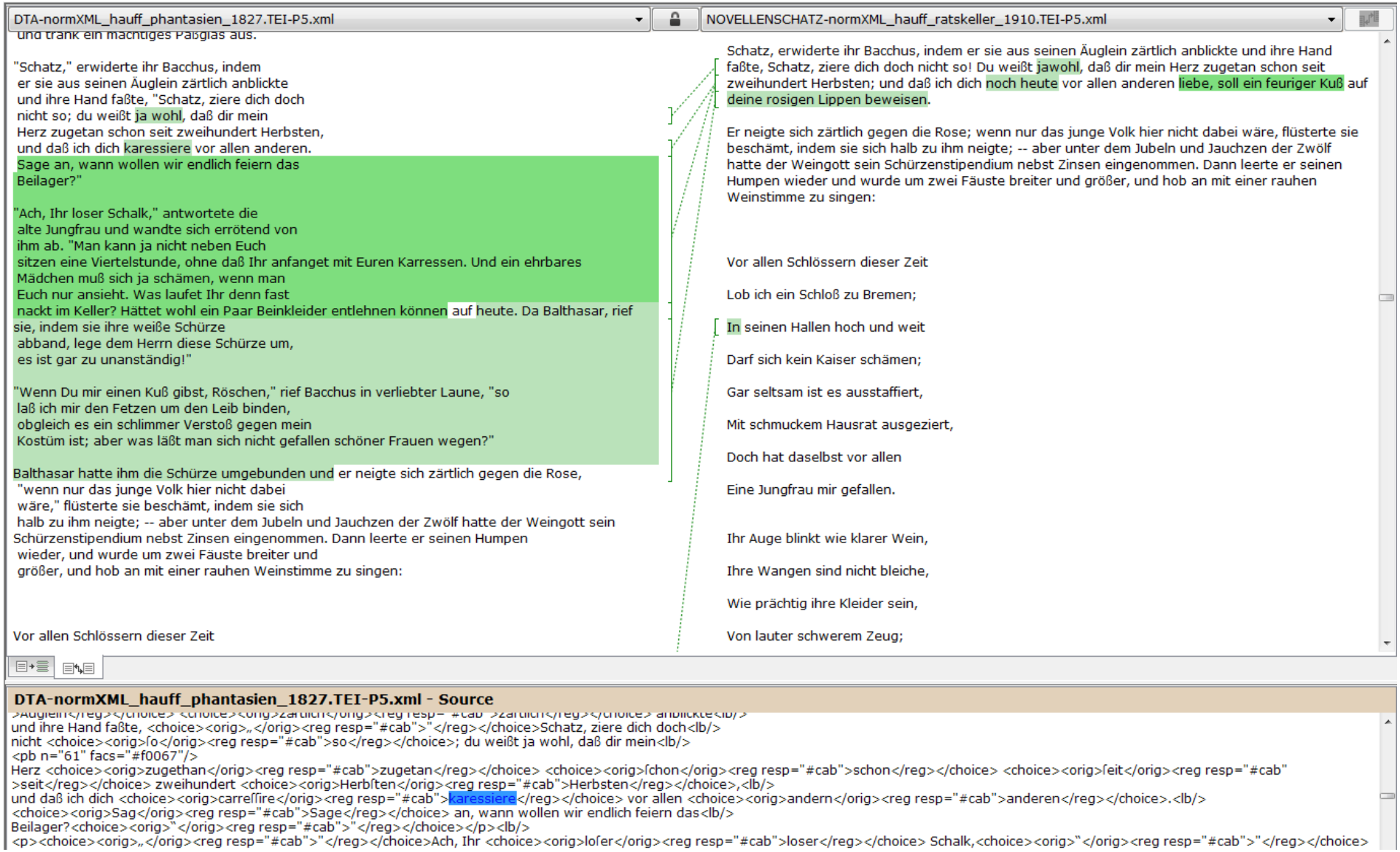
DTAQ – Tools

http://www.deutschestextarchiv.de/dtaq/tool

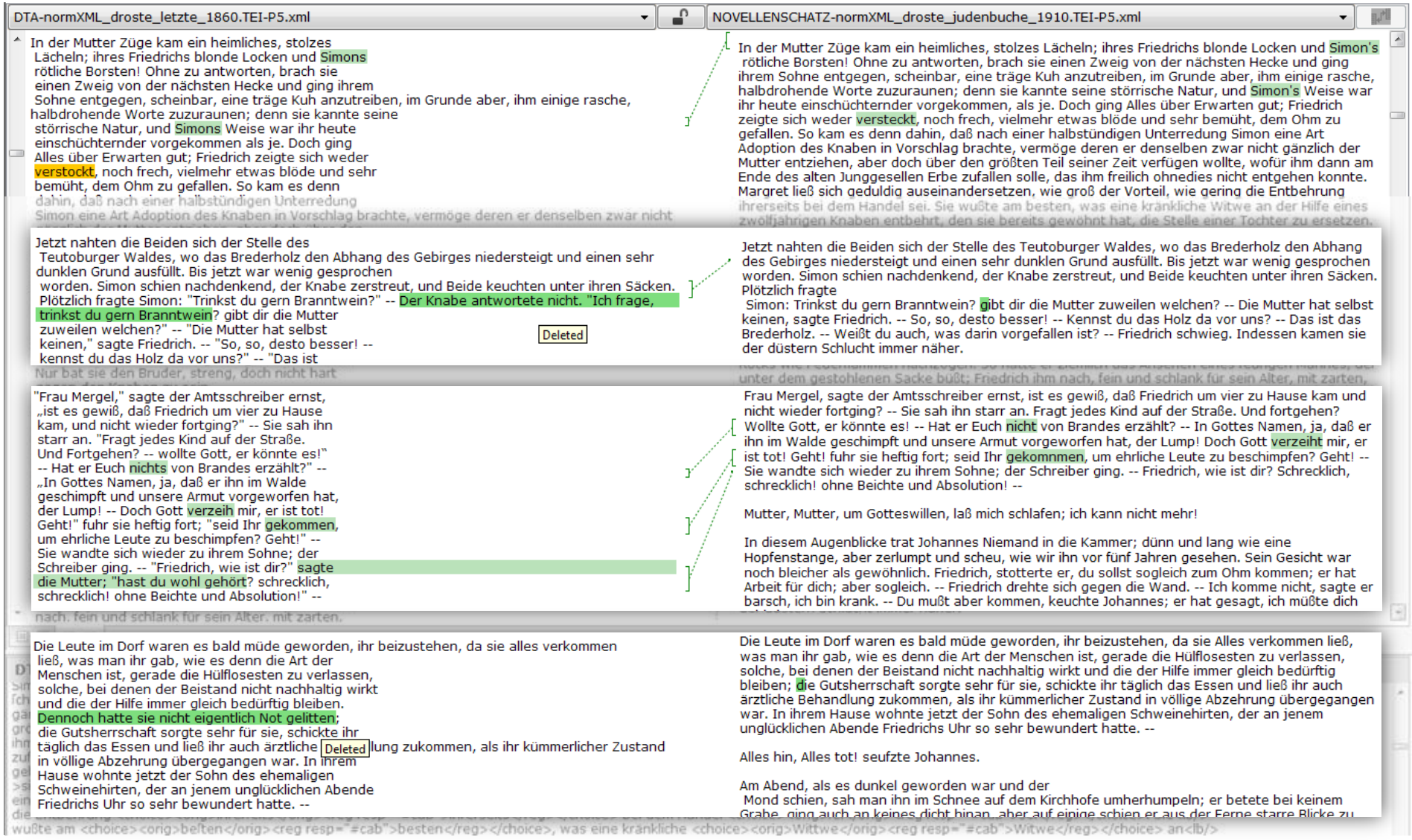
Kollation zweier Ausgaben mit juXta (Brentano)



Kollation zweier Ausgaben mit juXta (Hauff)



Kollation zweier Ausgaben mit juXta (Droste-Hülshof)



Korpus „Novellenschatz“ im DTA-Kontext (Ausblick)

- vorlagengetreue, strukturierte Transkription, standardisierte Annotation nach DTABf
  - vergleichsweise hoher Aufwand
  - ermöglicht Nutzung vieler Tools & Routinen:
    - computerlinguistische Analyse im DTA
    - (externe) Recherche- und Analysewerkzeuge
    - kollaborative Korrekturumgebung DTAQ
    - XML→HTML, XML→TCF, TEI-Header→CMDI usw.

Vielen Dank!

Kontakt: dta@bbaw.de, {thomas,wuerzner}@bbaw.de

Links

- DTA-Web: www.deutschestextarchiv.de
- DTAQ: www.deutschestextarchiv.de/dtaq
- DTABf: www.deutschestextarchiv.de/doku/basisformat
- DiaCollo: http://kaskade.dwds.de/dstar/dta/diacollo/
- BBAW-CLARIN-Zentrum: http://clarin.bbaw.de

Mehr

Publikationen: www.deutschestextarchiv.de/doku/publikationen

Blog: http://sprache.hypotheses.org

Twitter: @textarchiv