



Methoden zur effektiven Nachnutzung von Textressourcen aus dem Kontext wissenschaftlicher Projekte und Bibliotheken

Matthias Schulz, Susanne Haaf, Christian Thomas,
Bryan Jurish, Frank Wiegand, Alexander Geyken

dta@bbaw.de

<http://www.deutschestextarchiv.de>

Gliederung

- Prolog
- Fallbeispiel 1: Manuell erstellte Volltexte
- Fallbeispiel 2: Mithilfe von OCR-Verfahren erstellte Volltexte
- Fallbeispiel 3: Manuelle erfasste und TEI-kodierte Volltexte

in den Titeldaten
 im Korpus
 in der Dokumentation
 [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

Das Deutsche Textarchiv auf dem 1. Wikisource-Treffen in Leipzig

Vom 18. bis 20. Oktober 2013 fand in Leipzig das 1. Wikisource Arbeitstreffen statt. Das Deutsche Textarchiv war mit einem Beitrag zu Gast, in dem die bisherige Arbeit mit Wikisource-Texten im Rahmen des CLARIN-D Kurationsprojekts „Integration und Aufwertung historischer Textressourcen des 15.–19. Jahrhunderts in einer nachhaltigen CLARIN-Infrastruktur“ vorgestellt wurde. [mehr ...](#)

Vortrag: Historische Textkorpora für die Geistes- und Sozialwissenschaften. Das Beispiel „Deutsches Textarchiv“

Das DTA in Zahlen

- 1 170 Werke
- 374 291 digitalisierte Seiten
- 86 677 581 fortlaufende Wortformen
- 608 129 488 Zeichen (Unicode)
- 486 weitere Werke in DTAQ

Das DTA am 22. November 2013



Nachweis: Joseph Anton Koch, *Tiroler Landsturm 1809*

"Auf Granitbo Andreas Hofe Sandwirth vor geboren, und Granitfelsen s seine glorreich Schlachten zu Tyrols." Auf d erinnert Herm Alexander Ber

seinem Werk *Die Alpen in Natur- und Le* an den Tiroler Freiheitskämpfer Andreas am 22. November 1767 geboren wurde. Bemerkung Berlepschs geschieht freilich Rande einer Abhandlung über den Grani Ein eigener Bericht zu Leben und Taten Hofers findet sich in Johann Peter Hebe *Schatzkästlein des rheinischen Hausfre*



Berlepsch, Hermann Alexander: *Natur- und Lebensbildern.* Leipz



Hebel, Johann Peter: *Schatzkäs rheinischen Hausfreundes.* Tübir

Neue Werke im DTA

Verstreute Ressourcen

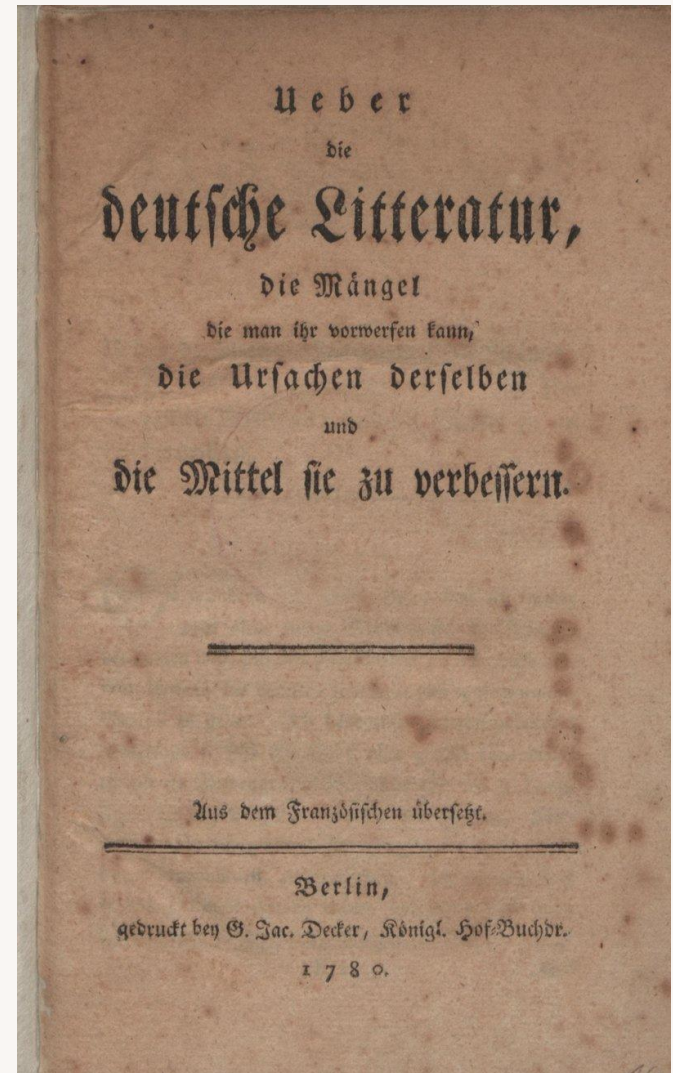
"Wir besitzen noch keine von der ganzen Nation gebilligte Sammlung, in der man alle Worte und Redensarten fände, nach denen man die Reinigkeit der Sprache sicher beurtheilen könnte."

Friedrich II.: Über die deutsche Literatur [...].

Aus dem Französischen übersetzt.

Berlin, 1780, S. 6. Online-Version:

<http://www.deutschestextarchiv.de/friedrich/literatur/1780/viewer/image/text/40/7/>



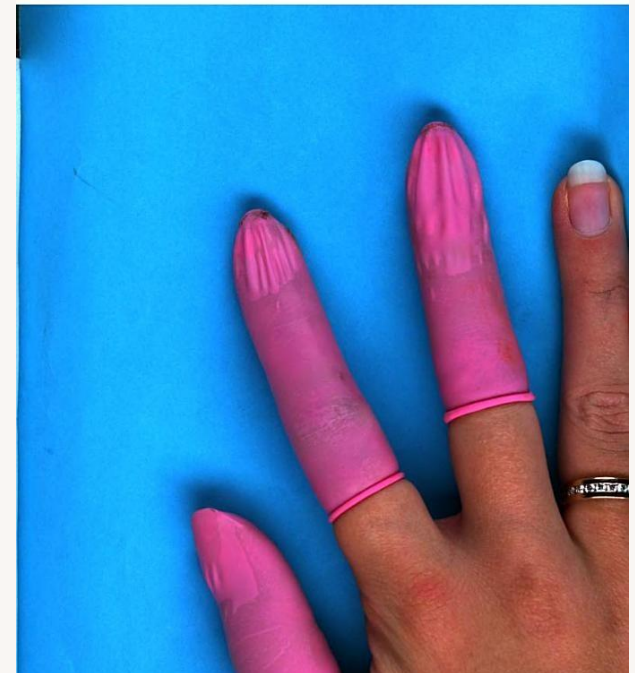
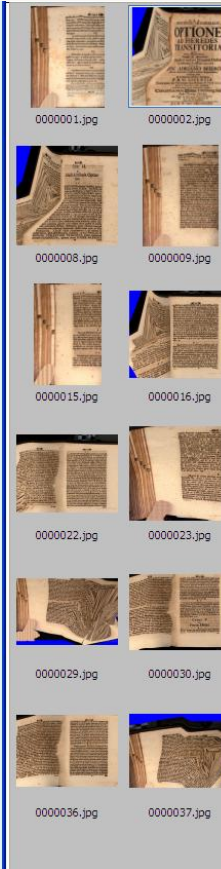
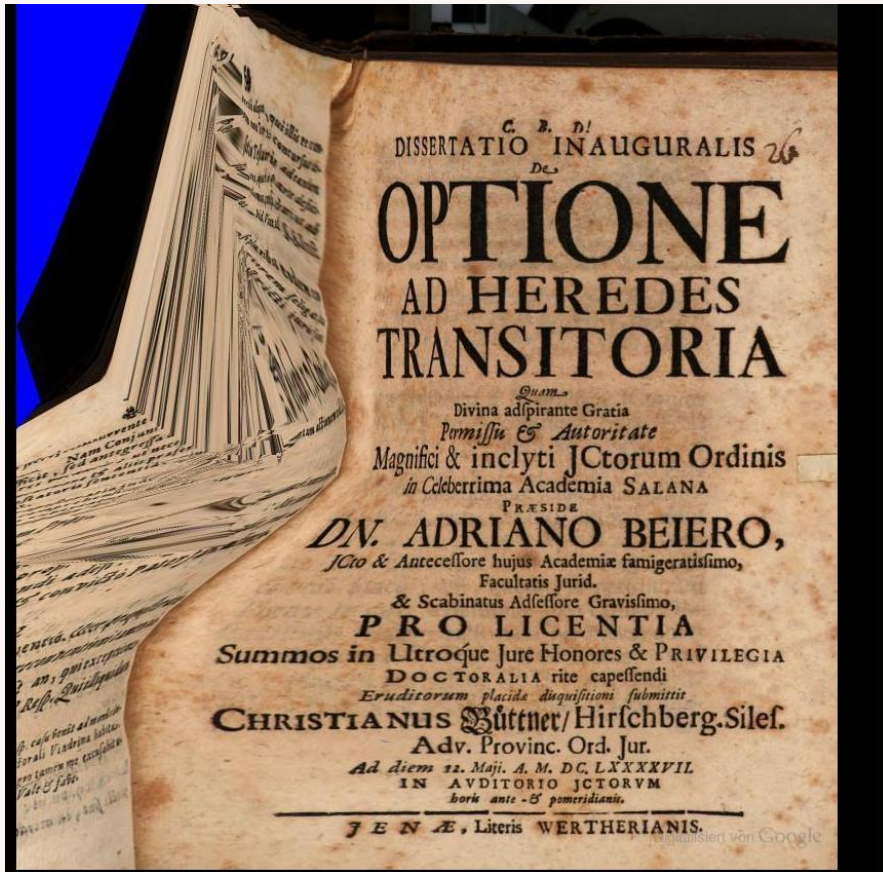
Nachweis: Berlin SBB-PK, Bibl. Diez oct. 8354

Defekte Vorlagen

<http://theartofgooglebooks.tumblr.com/>

BEAUTIFUL MISTAKES

The Accidental Art Of Google Books



Fehlerhafte Texte



Das tote Herz.

Mein Herz ist tot," einst eine Dame
Im Kreis intimer Freunde spricht,
„Die Liebe, ach, die wundersame,
Wird wieder es erwecken nicht.“

Aas tote **Kerz**.

Lein Herz ist tot," einst eine Dame

^ In, Areis intimer Freunde spricht,

[...]

F. C. Theodor Krüger: Was ich ersonnen, was ich erlebt! Gedichte. New York : Im Selbstverlage des Verfassers, 1897, S. 7, online-Version
<http://hdl.handle.net/2027/wu.89068036045>

Fehlerhafte Metadaten

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

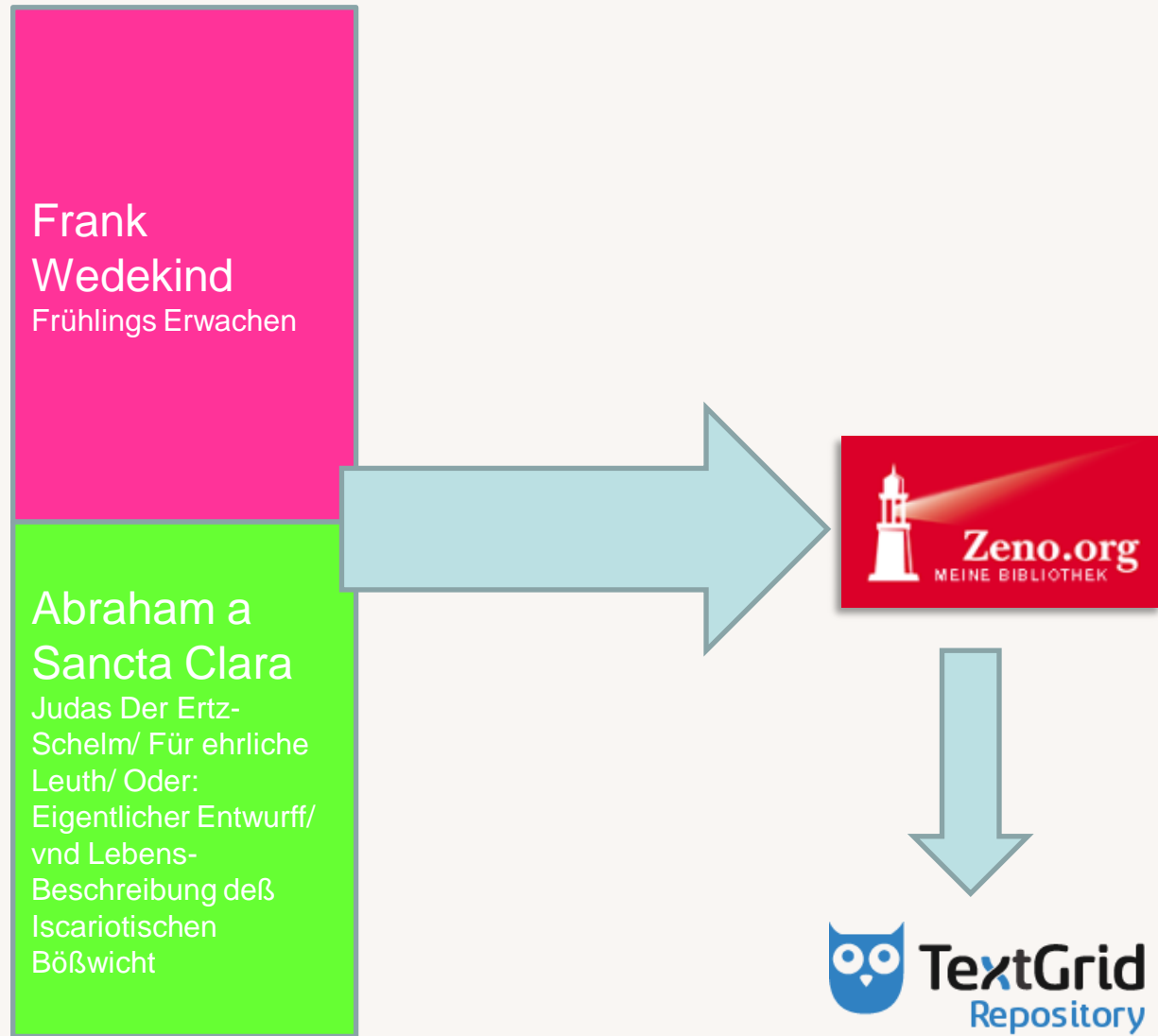
im Gegensatz

früherer Beleg "faktisch" in GoogleBooks: **1688**, <http://tinyurl.com/omynbsn>



The screenshot shows a Google search for "faktisch" with the "Bücher" (Books) tab selected. The search results display a book entry for "Schriften - Band 1; Band 3 - Seite 15" by Philipp Jakob Spener, published in 1688. The snippet of text from the book reads: "Spener hat keine zusammenhängende Ethik geschrieben; aber eine Untersuchung über Speners Ethik geriet bei der Rechenschaftsablage über ihre Quellen im Grunde **faktisch** zu einer Teileinführung in Speners 11) Zur Wichtigkeit einer ...". The word "faktisch" is highlighted in bold in the original image.

Zweifelhafte Vorlagen





Judas wird in der Insel Iscarioth / von
 dero er den Namen geschöpft / bey Hof / als ein
 Königlicher Prinz auferzogen / so aber bald das
 Gottlose Gemüth durch seinen Neyd an Tag
 geben.

Lasset sich doch noch reden / das gemeine Sprich-
 wort / wie grösser der Schelm / je besser das Glück /
 zumahlen diser Judas von den Meer-Wellen
 verschont worden / vnd so ohnverhofft zu diser Würde ge-
 langet / daß er als ein Königlicher Prinz ist auferzogen
 worden / den hat man in ein verguldrte Wiegen gelegt /
 da ihme doch der Sautrog hätte sollen die Herberg geben /
 den hat man in die zarteste Windelein eingefätscht / da doch
 dem Ohnflath die Zigeuner-Fetzen zu gut waren ; den hat

Judas wird in der Insel Iscarioth, von der er den Namen geschöpft, bei Hof als ein königlicher Prinz auferzogen, so aber bald das gottlose Gemüth durch seinen Neid an Tag geben.

Es **lässet** sich doch noch reden das gemeine Sprichwort: wie größer der Schelm, je besser das Glück, zumalen dieser Judas von den Meerwellen verschont worden und so **unverhofft** zu dieser Würde gelanget, daß er als ein königlicher Prinz ist auferzogen worden. Den hat man in eine **vergulde** Wiegen gelegt, da ihm doch der Sautrog hätte sollen die Herberg geben; den hat man in die zarteste Windelein eingefätscht, da doch dem Unflath die Zigeunerfetzen zu gut waren

[Transkription nach: Abraham a Sancta Clara: Judas der Erzschem für ehrliche Leuth, oder eigentlicher Entwurf und Lebensbeschreibung des Iscariotischen Böswicht. 7 Bände, in: Abraham a St. Clara's Sämmtliche Werke, Band 1, Passau: Friedrich Winkler, 1834–1836.]

Quellen:

1. Bildvorlage der Transkription im DTA

(Erstausgabe)

(http://www.deutschestextarchiv.de/dtaq/web/book/show/santa_judas01_1686)

2. Textgrid/zeno

<http://www.zeno.org/Literatur/M/Abraham+a+Sancta+Clara/Predigtliteratur/Judas+der+Erzschem/Erster+Band/Judas+wird+in+der+Insel+Iscarioth+auferzogen>

Zweifelhafte Vorlagen

Verabscheu**en**swürdiges] Verabscheu**ung**swürdiges

unverrichteter Sache] unverrichteter**s**ache

zu Muthe] zu**u**mute

eine **D**rainage in die Stirnhöhle appliciren zu lassen] [...] einen **V**entilator [...]

Ist's denn nicht der **s**tupendeste Beweis] Ist's denn nicht der **e**klatanteste Beweis

Der **e**inmalige Fehltritt könne noch zu ihrem Glücke führen] Der **e**hemalige [...]

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](http://nbn-resolving.org/urn:nbn:de:hbz:5:1-63868-p0071-9)

abgerufen am 17.09.2013.

Ziel

- Gewinnung oder Erstellung einer verlässlichen Transkription, bei:
- **größtmöglicher** Bewahrung des Vorlagentextes und gleichzeitiger Konzentration auf die **lexikalischen Gegebenheiten**

Gliederung

- Prolog
- **Fallbeispiel 1: Manuell erstellte Volltexte**
- Fallbeispiel 2: Mithilfe von OCR-Verfahren erstellte Volltexte
- Fallbeispiel 3: Manuelle erfasste und TEI-kodierte Volltexte

Fallbeispiel 1: Beispiel Wikisource



WIKISOURCE

Hauptseite
Systematischer Einstieg
Themenübersicht
Autorenindex
Zufällige Seite

Mitmachen
Letzte Änderungen
Neuer Artikel
Korrekturen des Monats
Gemeinschaftsportal
Skriptorium
Auskunft
Hilfe
Spenden

Drucken/exportieren

Werkzeuge

Sprachen

Benutzerkonto anlegen Anmelden

Quellentext [Index](#) [Diskussion](#)

Lesen [Bearbeiten](#) [Versionsgeschichte](#)

Suchen

Auf zum 1. Wikisourcetreffen vom 18.–20.10.2013 in Leipzig! Infos zur Anmeldung und zum Programm unter Wikisourcetreffen 2013

Unterhaltungen deutscher Ausgewanderten

[1-49]

Unterhaltungen deutscher Ausgewanderten.

In jenen unglücklichen Tagen, welche für Deutschland, für Europa, ja für die übrige Welt die traurigsten Folgen hatten, als das Heer der Franken durch eine übelverwahrte Lücke in unser Vaterland einbrach, verließ eine edle Familie ihre Besitzungen in jenen Gegenden und entfloh über den Rhein, um den Bedrängnissen zu entgehen, womit alle ausgezeichnete Personen bedrohet waren, denen man zum Verbrechen machte, daß sie sich ihrer Väter mit Freuden und Ehren erinnerten, und mancher Vortheile genossen, die ein wohlthätiger Vater seinen Kindern und Nachkommen so gern zu verschaffen wünschte.

Die Baronesse von C., eine Wittve in mittlern Jahren, erwieß sich auch jetzt auf dieser Flucht, wie sonst zu Hause, zum Troste ihrer Kinder, Verwandten und Freunde, entschlossen und thätig. In einer weiten Sphäre erzogen und durch mancherley Schicksale ausgebildet, war sie als eine treffliche Hausmutter bekannt, und jede Art von Geschäft erschien ihrem durchdringenden Geiste willkommen. Sie wünschte Vielen zu dienen und ihre ausgebreitete Bekanntschaft setzte sie in den Stand es zu thun. Nun mußte sie sich unerwartet als Führerin einer kleinen Caravane darstellen und wußte auch diese zu leiten, für sie zu sorgen und den guten Humor, wie er sich zeigte, in ihrem Kreise, auch mitten unter Bangigkeit und Noth zu unterhalten. Und wirklich stellte sich [1-50] bey unsern Flüchtlingen die gute Laune nicht selten ein, denn überraschende Vorfälle, neue Verhältnisse gaben den aufgespannten Gemüthern manchen Stoff zu Scherz und Lachen.

Bey der übereilten Flucht war das Betragen eines jeden charakteristisch und auffallend. Das eine ließ sich durch eine falsche Furcht, durch ein unzeitiges Schrecken hinreißen; das andere gab einer unnöthigen Sorge Raum, und alles, was dieser zu viel jener zu wenig that, jeder Fall wo sich Schwäche in Nachgiebigkeit oder Uebereilung zeigte, gab in der Folge Gelegenheit sich wechselseitig zu plagen und aufzuziehen, so daß dadurch diese traurige Zustände lustiger wurden, als eine vorsätzliche Lustreise ehemals hatte werden können.

Denn wie wir manchmal in der Comödie eine Zeitlang ohne über die absichtlichen Possen zu lachen, ernsthaft zuschauen können, dagegen aber sogleich ein lautes Gelächter entsteht, wenn in der Tragödie etwas Unschickliches vorkommt: so wird auch ein Unglück in der wirklichen Welt das die Menschen aus ihrer Fassung bringt gewöhnlich von lächerlichen, oft auf der Stelle, gewiß aber hinterdrein, belachten Umständen begleitet seyn.

Besonders mußte Fräulein *Luise*, die älteste Tochter der Baronesse, ein lebhaftes, heftiges und in guten Tagen herrisches Frauenzimmer sehr vieles leiden, da von ihr behauptet wurde, daß sie bey dem ersten Schrecken ganz aus der Fassung gerathen sey, in Zerstreung, ja in einer Art von völligen Abwesenheit die unnütze Sachen mit dem größten Ernste zum Aufpacken gebracht, [1-51] ja sogar einen alten Bedienten für ihren Bräutigam angesehen habe.

Sie vertheidigte sich aber so gut sie konnte, nur wollte sie keinen Scherz der sich auf ihren Bräutigam bezog, dulden, indem es ihr schon Leiden genug verursachte, ihn bey der allirten Armee, in täglicher Gefahr, zu wissen, und eine gewünschte Verbindung durch die allgemeine Zerrüttung aufgeschoben und vielleicht gar vereitelt zu sehen.

Ihr älterer Bruder *Friedrich*, ein entschloßner junger Mann, führte alles was die Mutter beschloß mit Ordnung und Genauigkeit aus, begleitete zu Pferde den Zug und war zugleich Courier, Wagenmeister und Wegweiser. Der Lehrer des jüngern, hoffnungsvollen Sohnes, ein wohl unterrichteter Mann, leistete der Baronesse im Wagen Gesellschaft; Vetter *Karl* fuhr mit einem alten Geistlichen, der, als Hausfreund, schon lange der Familie unentbehrlich geworden war, mit einer älteren und jüngeren Verwandten in einem nachfolgenden Wagen. Kammermädchen und Kammerdiener folgten in Halbehaisen, und einige schwerbepackte Brankards, die auf mehr als einer Station zurückbleiben mußten, schlossen den Zug.

Ungern hatte, wie man leicht denken kann, die ganze Gesellschaft ihre Wohnungen verlassen, aber Vetter *Karl* entfernte sich mit doppeltem Widerwillen von dem einsamen Rheinufer, nicht daß er etwa eine Geliebte daselbst zurücklassen hätte, wie man nach seiner Jugend, seiner guten Gestalt und seiner leidenschaftlichen

Textdaten	
Autor:	Johann Wolfgang von Goethe
Titel:	Unterhaltungen deutscher Ausgewanderten
aus:	Die Horen 1795: 1. Band, 1. Stück S. 49–78; 2. Stück S. 1–28; 2. Band, 4. Stück S. 41–67; 3. Band, 7. Stück S. 50–76; 9. Stück S. 45–52; 4. Band, Zehntes Stück S. 108–152 "Mährchen"
Herausgeber:	Friedrich Schiller
Auflage:	1. Auflage
Erscheinungsdatum:	1795
Verlag:	Cotta
Erscheinungsort:	Tübingen
Quelle:	Hathitrust und Commons
Artikel in der Wikipedia	

Bearbeitungsstand	
korrigiert	
Dieser Text wurde anhand der angegebenen Quelle einmal Korrektur gelesen. Die Schreibweise sollte dem Originaltext folgen. Es ist noch ein weiterer Korrekturdurchgang nötig.	
Um eine Seite zu bearbeiten, brauchst du nur auf die entsprechende [Seitenzahl] zu klicken. Weitere Informationen findest du	

URL: [http://de.wikisource.org/wiki/Unterhaltungen deutscher Ausgewanderten](http://de.wikisource.org/wiki/Unterhaltungen_deutscher_Ausgewanderten)

DTAE – Import aus Wikisource

Schritt 1: Datenimport vorbereiten

Bitte geben Sie den Seitentitel des Textes bei Wikisource ein. Bitte beachten Sie, dass das ausgewählte Werk keine Teilseite eines übergeordneten Werkes ist.

TEI aus o. g. Wikisource-Seite erstellen
 TEI aus Seiten für o. g. Index erstellen



Hole Informationen von Wikisource. Dieser Vorgang kann einen Moment dauern ...

Metadaten-Import ,live‘

Metadaten

- Dateien - Download
 Bilddateien · HTML-Dateien · generiertes TEI
Vorgesehener DTA-Verzeichnisname:

skriptbasierte Textdaten-Konvertierung nach DTABf

Textvorlage

Autor:
Titel:
Untertitel:
aus:
Herausgeber:
Ausgabe:
Verlag:
Ort:
Jahr:
vorw. Schrifttype: Fraktur Antiqua keine Angabe

Texttranskription

Es gelten folgende **Editionsrichtlinien** (HTML). Angaben zur Formatierung können hierbei entfernt werden.

Folgende Lizenz gilt für den redaktionellen Teil der Transkription:

Bildvorlage

Als Quelle für die Bilddigitalisate wurde folgendes ermittelt:

- [Hathitrust](#)
- [Commons](#)

weiter zur Erstellung des TEI-Headers



Erster Act.

Anmuthige Gegend,

Erster Act.

Anmuthige Gegend.

Faust auf blumigen Rasen gebettet, ermüdet, unruhig, schlafsuchend.
Dämmerung.
Geister-Kreis schwebend bewegt, anmuthige kleine Gestalten.

Wikisource Markup:
Information Ebene = semantische
Informationen sind **implizit**,
typographische Informationen explizit

```
{{LineCenterSize|130|25|'''Erster Act.'''}}
```

```
{{Linie}}
```

```
{{LineCenterSize|130|25|'''Anmuthige Gegend.'''}}
```

```
{{Linie}}
```

```
{{Center|'''Faust''' auf blumigen Rasen gebettet,  
ermüdet, unruhig, schlafsuchend.'''}}
```

```
{{Center|'''Dämmerung.'''}}
```

```
{{Center|'''Geister-Kreis''' schwebend bewegt,  
anmuthige kleine Gestalten.'''}}
```

"big" + "bold" + "centred"

"horizontal rule"

"bold" ± "italics" + "centred"

Erster Act.

Anmuthige Gegend.

DTABf

WE USE TEI

DTABf:
Information Ebene =
semantische **und** typographische
Informationen sind **explizit** enthalten

= Start of text `<body>`

= an element marking a section
(rendered as hr)

= Start `<div>`, type "act"

= Start `<div>`, type "scene"

= `<head>`line (centred)

= `<stage>` direction (centred)
(typeset in spaced letters)

Erster Act.

Anmuthige Gegend.

Faust auf blumigen Rasen gebettet, ermüdet, unruhig, schlafsuchend.

Dämmerung.

Geister-Kreis schwebend bewegt, anmuthige kleine Gestalten.

```
<facets facs="#f0010" n="[3]"/>
<body>
<milestone rendition="#hr" unit="section"/><lb/>
<div n="1" type="act">
<head><hi rendition="#g">Erster Act.</hi>.</head><lb/>
<milestone rendition="#hr" unit="section"/><lb/>
<div n="2" type="scene">
<head><hi rendition="#g">Anmuthige Gegend.</hi>.</head><lb/>
<stage><hi rendition="#c"><hi rendition="#g">Faust.</hi> auf b
suchend.</hi></stage>
<stage><hi rendition="#c"><hi rendition="#g">Dämmerung.</hi></stage>
<stage><hi rendition="#c"><hi rendition="#g">Geister-Kreis/</hi></stage>
stalten.</hi></stage>
```


Gliederung

- Prolog
- Fallbeispiel 1: Manuell erstellte Volltexte
- **Fallbeispiel 2: Mithilfe von OCR-Verfahren erstellte Volltexte**
- Fallbeispiel 3: Manuelle erfasste und TEI-kodierte Volltexte

rakter wie unsere „Schnaderhüpfel“ in Tirol und Kärnten. In den kleinen unüberschriebenen Gedichten der Abteilung „Liebeschiller,“ Gefühlsafforden, kurzen Stimmungslauten, hat er das beste Zeugnis für seine rein lyrische Begabung abgelegt. Als „Anhang“ bietet er drei Balladen, von denen die eine: „Byzantinisches Triumphbild“ ein herbes Pathos und eine kräftige Gestaltungskraft bekundet. Es genüge hier, darauf hinzuweisen.

<abb>



Die Smaragdinsel.

<abb>



Das neunzehnte Jahrhundert ist das Jahrhundert des Nationalitätsprinzips. Alle Bewegungen und Umwälzungen, die in Europa in neuerer Zeit stattgefunden haben, haben als Grundlage das Streben nach Durchführung dieses Prinzips gehabt. Wenn wir daher von den Vorgängen in Irland lesen, von dem hartnäckigen Kampfe der Iren um Selbständigkeit, so sind wir geneigt, auch hier das Walten des mächtigen Geistes zu erkennen, der die Menschen heißt dem Drange des Blutes folgen und sich Verhältnissen widersetzen, welche die Natur verachten. *)

Die Engländer sind germanischen Ursprunges, die Iren keltischen. Aber auch die Hochlandschotten und die Walliser gehören der keltischen Familie an und haben die keltische Sprache treuer bewahrt als die Iren. Dennoch hören wir nirgends, daß die Walliser ein eignes Parlament beanspruchten oder dem englischen Nachbar feindlich wären. Der Walliser, auch in den engen Thälern, wo er seine Sprache frei von englischen Einflüssen bewahrt hat, ist stolz, ein Unterthan der Königin Viktoria zu sein, und gedeiht unter dem Schutze des britischen Löwen.

Anders die Iren. In Irland ist eine notleidende, unglückliche Bevölkerung, ohne Besitz, von oben bedrückt und von seinen Führern zur Gewaltthat dem Drucke gegenüber aufgereizt.

Wales wurde im Jahre 1282 von Edward I. erobert, nachdem Irland schon im Jahre 1169 unterworfen und im nördlichen Teile besiedelt worden war. Zeit genug haben die beiden Länder gehabt, sich an das Neue zu gewöhnen. Wenn sich Irland noch heute gegen England sträubt und die Iren

*) Dieser Aufsatz wird unseren Lesern als Ergänzung zu dem im 19. Hefte: „Erinnerungen aus Irland“ willkommen sein.

rakter wie unsere „Schnaderhüpfel“ in Tirol und Kärnten. In den kleinen unüberschriebenen Gedichten der Abteilung „Liebeschiller,“ Gefühlsafforden, kurzen Stimmungslauten, hat er das beste Zeugnis für seine rein lyrische Begabung abgelegt. Als „Anhang“ bietet er drei Balladen, von denen die eine: „Byzantinisches Triumphbild“ ein herbes Pathos und eine kräftige Gestaltungskraft bekundet. Es genüge hier, darauf hinzuweisen.

<abb>



d1

Die Smaragdinsel.



Das neunzehnte Jahrhundert ist das Jahrhundert des Nationalitätsprinzips. Alle Bewegungen und Umwälzungen, die in Europa in neuerer Zeit stattgefunden haben, haben als Grundlage das Streben nach Durchführung dieses Prinzips gehabt. Wenn wir daher von den Vorgängen in Irland lesen, von dem hartnäckigen Kampfe der Iren um Selbständigkeit, so sind wir geneigt, auch hier das Walten des mächtigen Geistes zu erkennen, der die Menschen heißt dem Drange des Blutes folgen und sich Verhältnissen widersetzen, welche die Natur verachten. *)

Die Engländer sind germanischen Ursprunges, die Iren keltischen. Aber auch die Hochlandschotten und die Walliser gehören der keltischen Familie an und haben die keltische Sprache treuer bewahrt als die Iren. Dennoch hören wir nirgends, daß die Walliser ein eignes Parlament beanspruchten oder dem englischen Nachbar feindlich wären. Der Walliser, auch in den engen Thälern, wo er seine Sprache frei von englischen Einflüssen bewahrt hat, ist stolz, ein Unterthan der Königin Viktoria zu sein, und gedeiht unter dem Schutze des britischen Löwen.

Anders die Iren. In Irland ist eine notleidende, unglückliche Bevölkerung, ohne Besitz, von oben bedrückt und von seinen Führern zur Gewaltthat dem Drucke gegenüber aufgereizt.

Wales wurde im Jahre 1282 von Edward I. erobert, nachdem Irland schon im Jahre 1169 unterworfen und im nördlichen Teile besiedelt worden war. Zeit genug haben die beiden Länder gehabt, sich an das Neue zu gewöhnen. Wenn sich Irland noch heute gegen England sträubt und die Iren

fn

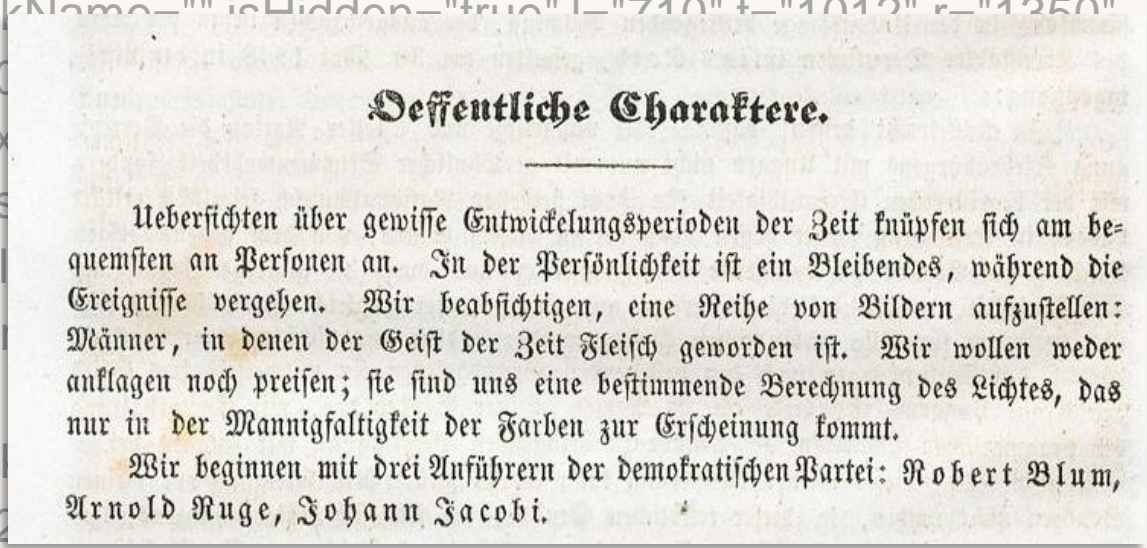
*) Dieser Aufsatz wird unseren Lesern als Ergänzung zu dem im 19. Hefte: „Erinnerungen aus Irland“ willkommen sein.

Ausgangspunkt: ABBYY-OCR-Text

```
<charParams l="669" t="2025" r="702" b="2059" characterHeight="24" [...]  
  charConfidence="98" serifProbability="255">D</charParams>  
<charParams l="706" t="2035" r="722" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">a</charParams>  
<charParams l="726" t="2035" r="742" b="2068" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">g</charParams>  
<charParams l="745" t="2035" r="763" b="2059" suspicious="true" characterHeight="24" [...]  
  charConfidence="98" serifProbability="255">u</charParams>  
<charParams l="768" t="2035" r="780" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">e</charParams>  
<charParams l="781" t="2036" r="796" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">r</charParams>  
<charParams l="798" t="2036" r="812" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">r</charParams>  
<charParams l="815" t="2036" r="830" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">o</charParams>  
<charParams l="833" t="2029" r="844" b="2059" characterHeight="24" [...]  
  charConfidence="100" serifProbability="255">t</charParams>  
<charParams l="846" t="2033" r="862" b="2068" suspicious="true" [...] characterHeight="24"  
  charConfidence="92" serifProbability="255">y</charParams>  
<charParams l="866" t="2033" r="884" b="2068" characterHeight="24" [...]  
  charConfidence="96" serifProbability="255">p</charParams>
```

Ausgangspunkt: ABBYY-OCR-Text

```
<block blockType="Text" blockName="" isHidden="true" l="710" t="1012" r="1250"
b="1077"><region><rect l="710"
b="1077"></rect></region><text>
<par align="Justified"><line bas
b="1071"><formatting lang="Old
Deffentlichs Charaktere.</fo
</block>
```



```
<block blockType="Text" blockName="" isHidden="true" l="302"
b="2420"><region><rect l="302"
b="2420"></rect></region><text><par align="Justified" rightIndent="2"
startIndent="87" lineSpacing="55"><line baseline l="396" t="1177" r="1747"
b="1223"><formatting lang="OldGerman" ff="Arial" fs="10.">
```


Uebersichten über gewisse Entwicklungsperioden der Zeit knüpfen sich am bequemsten an Personen an. In der Persönlichkeit ist ein Bleibendes, während die

```
</block>
```



Deffentliche Charaktere.

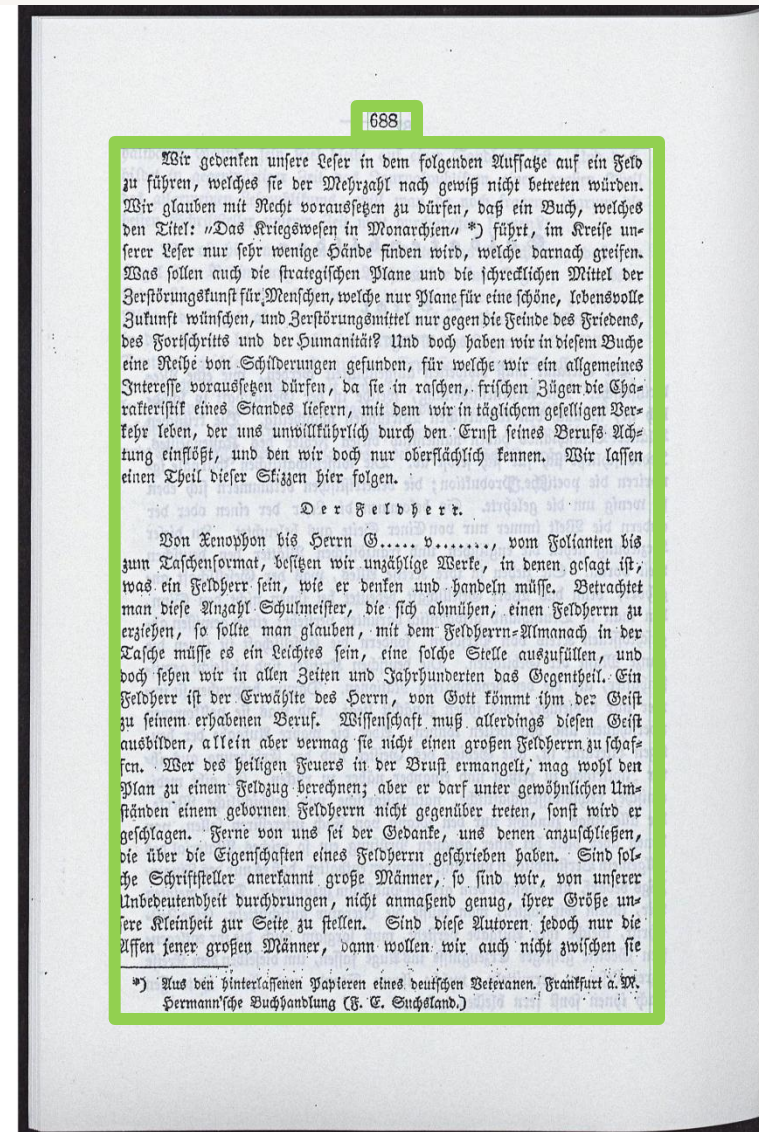
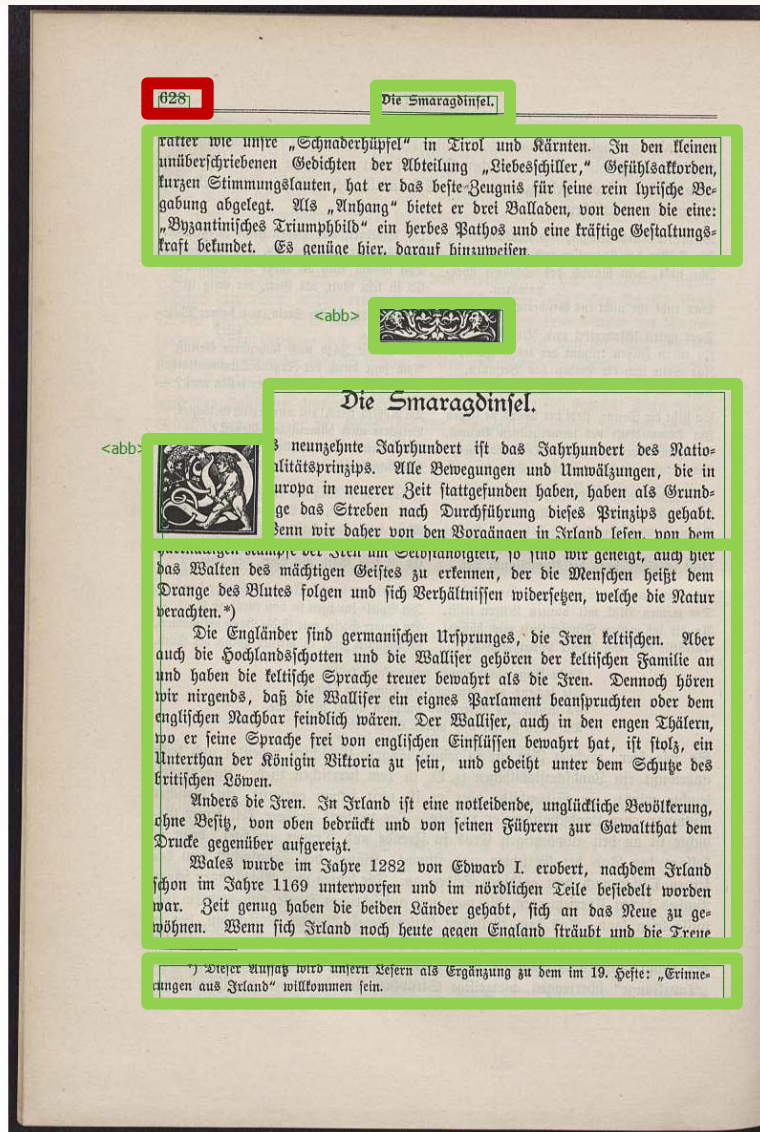
```
<par align="Justified">  
  <line baseline="1061" l="716" t="1018" r="1344" b="1071"><formatting lang="OldGerman"  
  ff="Arial" fs="13.">Deffentlichs Charaktere.</formatting></line>  
</par>
```



I. Robert Blum.

```
<par align="Center" rightIndent="2">  
  <line baseline="1710" l="818" t="1672" r="1238" b="1712"><formatting  
  lang="OldGerman" ff="Arial" fs="10.">I. Robert Blum.</formatting></line>  
</par>
```

ABBYY-OCR-Zonen (<block/>-Tags)



Nachbearbeitung mit DTA-Zoning Tool

The screenshot displays the DTA-Zoning Tool interface. On the left, a historical document titled "Die Smaragdinsel" is shown. The text describes the 19th century and the relationship between England and Ireland. A small illustration of a dragon is visible in the top left corner of the document. The tool's configuration window, titled "ZOT - Makrostruktur festlegen", is overlaid on the right side of the document. This window contains several panels for defining document structure:

- Text:** Includes options for "reiner Text", "Spalte", and "unbekannter Text".
- Text - Strukturmerkmale:** Includes "Seitenzahl", "Bogensignatur" (with a sample "©Müller, 1844 h. 1849/1850 3"), "Kustode", "lebende Kolummentitel" (with a sample "nach dem verschiedenen Charakter Herr Bester. 27. Manke bei Wert und in Mittel einander ausgeglichen, so sagt er von ihnen, daß sie in den Gärten der Gärten noch ganz Blig werden."), "Fußnote", "laufende Fußnote", and "Endnote".
- Marginalien:** Includes "Marginalie links" and "Marginalie rechts".
- Überschriften:** Includes options for "H1 Überschrift D1" through "H5 Überschrift D5", "Überschrift D6", "Überschrift D7", "Überschrift D8", and "Überschrift D - ohne Hierarchie".
- Textsorten:** Includes "Gedicht", "Gedichttitel", "Strophe eines Gedichts", "Liste", "Inhaltsverzeichnis", "Front", and "Back".
- grafische Elemente (nicht abtippen):** Includes "Bild", "Tabelle oder Übersichten", "Aufistung", "Formel", "horizontale Linie", and "Spezialbild - keine OCR-Bereich".
- China Zot - Spezialtypen:** Includes "Figurenaufstellung im Drama", "Sprecher im Drama", "Bühnenanweisung", "Anrede im Brief", "Abschließender Gruß im Brief", "Bildunterschrift", "Zitat", "Widmung", "Kapitelteaser", and "Briefdatum".

At the bottom of the document, a footnote reads: "*) Dieser Aufsatz wird unsern Lesern als Ergänzung zu dem im 19. Hefte: „Erinnerungen aus Irland“ willkommen sein."

Die Smaragdinsel.

d1



Das neunzehnte Jahrhundert ist das Jahrhundert des Nationalitätsprinzips. Alle Bewegungen und Umwälzungen, die in Europa in neuerer Zeit stattgefunden haben, haben als Grundlage das Streben nach Durchführung dieses Prinzips gehabt. Wenn wir daher von den Vorgängen in Irland lesen, von dem hartnäckigen Kampfe der Iren um Selbständigkeit, so sind wir geneigt, auch hier das Walten des mächtigen Geistes zu erkennen, der die Menschen heißt dem Drange des Blutes folgen und sich Verhältnissen widersetzen, welche die Natur verachten.*)

Die Engländer sind germanischen Ursprunges, die Iren keltischen. Aber auch die Hochlandsschotten und die Walliser gehören der keltischen Familie an und haben die keltische Sprache treuer bewahrt als die Iren. Dennoch hören wir nirgends, daß die Walliser ein eignes Parlament beanspruchten oder dem englischen Nachbar feindlich wären. Der Walliser, auch in den engen Thälern, wo er seine Sprache frei von englischen Einflüssen bewahrt hat, ist stolz, ein Unterthan der Königin Viktoria zu sein, und gedeiht unter dem Schutze des britischen Löwen.

Anderä die Iren. In Irland ist eine notleidende, unglückliche Bevölkerung, ohne Besitz, von oben bedrückt und von seinen Führern zur Gewaltthat dem Drucke gegenüber aufgereizt.

Wales wurde im Jahre 1282 von Edward I. erobert, nachdem Irland schon im Jahre 1169 unterworfen und im nördlichen Teile besiedelt worden war. Zeit genug haben die beiden Länder gehabt, sich an das Neue zu gewöhnen. Wenn sich Irland noch heute gegen England sträubt und die Treue

*) Dieser Aufsatz wird unsern Lesern als Ergänzung zu dem im 19. Hefte: „Erinnerungen aus Irland“ willkommen sein.

fn

manuelle Änderungen:

- Zonengrenzen korrigiert
- Schmuckinitiale als Teil des Textes markiert
- Textblöcke hinsichtlich ihrer semantischen Funktion differenziert (Seitenzahl, Kolummentitel, Abbildung/Buchschmuck, Überschrift (Ebene 1), Fußnote/Anmerkung, unmarkiert=Fließtext)

Konvertierung des strukturierter OCR-Rohtext

```
<d1> Deffentlichs Charaktere. </d1>  
<hr/>
```

```
<p>Uebersichten über gewisse Entwicklungsperioden der Zeit knüpfen sich am be-  
quemsten an Personen an. In der Persönlichkeit ist ein Bleibendes, während die  
Ereignisse vergehen. Wir beabsichtigen, eine Reihe von Bildern aufzustellen: [...]  
nur in der Mannigfaltigkeit der Farben zur Erscheinung kommt.</p>
```



```
<div n="1"> <head>Deffentlichs Charaktere.</head><lb />  
<milestone unit="section" rendition="#hr"/>
```

```
<p>Uebersichten über gewisse Entwicklungsperioden der Zeit knüpfen sich am be-<lb />  
quemsten an Personen an. In der Persönlichkeit ist ein Bleibendes, während die<lb />  
Ereignisse vergehen. Wir beabsichtigen, eine Reihe von Bildern aufzustellen:<lb />[...]  
nur in der Mannigfaltigkeit der Farben zur Erscheinung kommt.</p><lb /></div>
```

Gliederung

- Prolog
- Fallbeispiel 1: Manuell erstellte Volltexte
- Fallbeispiel 2: Mithilfe von OCR-Verfahren erstellte Volltexte
- **Fallbeispiel 3: Manuelle erfasste und TEI-kodierte Volltexte**



Quote this page



Transkription

[Braunschweig-Wolfenbüttel, Herzog Heinrich Julius von]: Landtags Abschiedt So zwischen dem Hochwürdigen/ Durchleuchtigen ... Herrn Heinrichen Julio/ Postulirtem Bischoffen zu Halberstadt/ unnd Hertzogen zu Braunschweig und Lüneburgk/ [et]c. Und S.F.G. Landtschafft des Fürstenthumbs Braunschweig Wolffebüttelschen theils zu Saltzdalum am Dritten Junii Anno 1597. auffgerichtet.

[Inhaltsverzeichnis]

|| [ID00001]

Landtags Abschiedt So zwischen dem Hochwürdigen / Durchleuchtigen / Hochgebornen Fürsten vnd Herrn / Herrn Heinrichen Julio / Postulirtem Bischoffen zu Halberstadt / vnnd Hertzogen zu Braunschweig vnd Lüneburgk / etc. ↑

Vnd S. F. G. Landtschafft des Fürstenthumbs Braunschweig Wolffebüttelschen theils zu Saltzdalum am Dritten Junij Anno 1597. auffgerichtet. ↑

Wolffenbüttel Durch Johan: Stangen / Anno 1604. ↑

|| [ID00002]



Richtlinien zur Texterfassung

4.4 r-Grapheme

Das sog. runde r wird mittels seiner entsprechenden Unicode-Entität (ꝛ, U+A75B, LATIN SMALL LETTER R ROTUNDA) abgebildet. Es findet sich häufig in Zusammenhang mit dem heute gebräuchlichen r oder als et-Substituent in Abkürzungen für et cetera.

Beispiel (rundes r als heutiges r und als et):

Vorlage	Transkription
	Herzꝛ#x75B;
	ꝛ#x75B;c.

4.5 Ligaturen

Vokalische Ligaturen werden grundsätzlich realisiert:

Vorlage	Zeichen	Entität	Beschreibung
	æ	U+00E6	LATIN SMALL LETTER AE
	œ	U+0153	LATIN SMALL LIGATURE OE

Konsonantische Ligaturen (tz, ct, ts, ff etc.) sowie die Ligatur ij werden dagegen grundsätzlich aufgespalten.

4.6 Umlaute

Umlaute werden entsprechend der Vorlage transkribiert, d. h. die Umlaute in den heute gebräuchlichen Formen ä, ö, ü, ä, ø, u werden von solchen, die durch ein hochgestelltes e (U+0364, COMBINING LATIN SMALL LETTER E) über Vokal gekennzeichnet sind, unterschieden (z. B. us#x0364;).

4.7 Kürzungsstriche

Kürzungsstriche (Balken oder geschlängelte Linie über Buchstaben als Substituenten für ausgelassene Zeichen, Nasalstrich, Geminationsstrich) werden mittels des Zeichens U+0303, COMBINING TILDE transkribiert.

Vorlage	Transkription
	from̄#x0303;en
	Un̄#x0303; macht

4.8 Diakritika

Diakritika werden nach Möglichkeit mittels Unicode-Entitäten realisiert, z. B. das hochgestellte o (U+0366, COMBINING LATIN SMALL LETTER O) über u bzw. U, das e-Cedille (ç, U+00E7, LATIN SMALL LETTER C WITH CEDILLA), die e caudata (e, U+0119, LATIN SMALL LETTER E WITH OGONEK) in der Bedeutung ae oder das e mit Trema (ë, U+00EB, LATIN SMALL LETTER E WITH DIAERESIS).

Weibern hochgehaltenen Aberglauben.

Das 3. Capitel.

Wenn man etwas von Wäſche 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-Hauffen/ denn ich habe es ſelbſt oft probiren müſſen/ wenn ich meine Wäſche auff einer Seite eingewäſcht gehabt/ ſo habe ich zuweilen Hembd und Halſtuch umgewendet/ und hernach auffſ neue darinnen gepranget/ als wie ein Bauer-Bräutigam. Und kan ich euch mit tauſend Eyden attetiren/ daß ich zu ſolcher Zeit niemahls bin beſchrien worden. Aber ſagt mir doch/ ihr guten Weiber: (denn ich zurne nicht mit euch/ ſondern habe nur einen Unglauben an euren Wercken) Wie gehet es zu/ daß man in verkehrter Wäſche nicht beſchrien werden kan? Ich will euch ſagen was ich dencke/ und weil ihr doch davor haltet/ daß man eine Sache glauben müſſe/ wann es helffen ſoll/ ſo will ich auch glauben/

Richtlinien zur Nutzung von Texterfassungen



Weibern hochgehaltenen Aberglauben.

Das 3. Capitel.

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

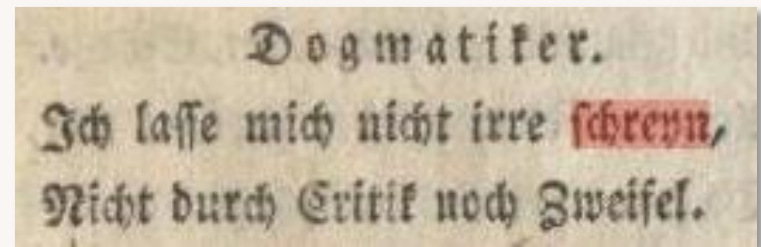
Das trifft gewiß ein / und gestehe ichs selbst.
Hey! Victoria! ihr Weiber habt recht; in
diesem Punct stehe ich euch bey biß an Scheiter-

- zeichengenau nach Vorlage
- keine Modernisierungen; keine Normalisierungen
- „Ich lasse mich nicht irre schreyen“

schreyen → schreyen ☺

schreyen → schreyen ☹

schreyen → schreien ☹



goethe_faust01_1808?p=293

Richtlinien zur TEI/XML-Annotation

Level 1: notwendig: `<cb/>`, `<list>`, `<lg>`, `<note>`

Level 2: empfohlen: `<choice>`, `<fw>`, `<lb/>`

Level 3: fakultativ: `<foreign>`, `<persName>`

Level 4: unzulässig: `<div1>`

- unterschiedliche editorische Erschließungstiefen
 - DTA-Kernkorpus: bis Level 2
- DTABf bietet Alternativen für unzulässige Elemente



Seitenzahl
Kolummentitel

Richtlinien zur TEI/XML- Annotation

Morgen = Gedanken

Gedicht

Titel

I 7 6 I.

Datum

horizontale Linie

Der Morgen dreht sein heitres Angesichte

Verse

Uns lächelnd zu, und weckt mit sanftem Lichte

Die Creaturen an den Tag hervor!

Der Sperling schwazt, die muntern Hähne krähen

Den Lobgesang, und aller Augen sehen,

Zu Gott, der sie ernährt, empor.

Strophen

Auch ich bin wach, und meinem ersten Blicke

Befehl ich, daß er Dank zum Himmel schicke

Für diese Ruh, für diese sanfte Nacht!

Es ist ein Gott, der diese Welt regieret,

Der aus dem Staub mich wunderbar geführet,

Und der mir Freud und Freunde macht!

Bogensignatur

Formal
Inhaltlich



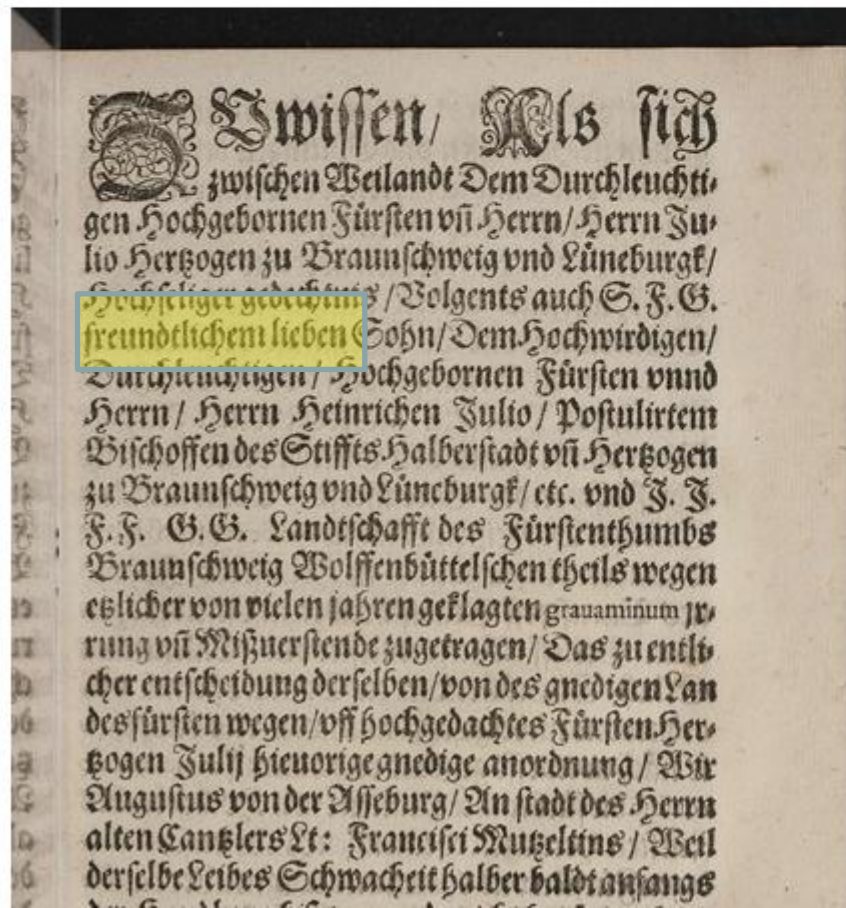
braunschweig_abschiedt_1604 (HAB)

offe
Stand

Bild: 0003 < vorherige Seite

nächste S

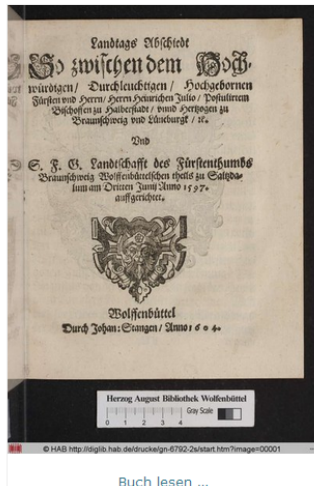
+ - 1:1 fit 36%



ZVwissen / Als sich zwischen Weilandt Dem Durchleuchtigen Hochgebornen Fürsten vn̄ Herrn / Herrn Julio Herzogen zu Braunschweig vnd Lüneburgk / Hochseliger gedächtnis / Volgents auch S. F. G. freundlichem lieben Sohn / Dem Hochwirdigen / Durchleuchtigen / Hochgebornen Fürsten vnn̄d Herrn / Herrn Heinrichen Julio / Postulirtem Bischoffen des Stiffts Halberstadt vn̄d Herzogen zu Braunschweig vnd Lüneburgk / etc. vnd J. J. F. G. Landtschafft des Fürstenthumbs Braunschweig Wolfenbüttelschen theils wegen eslicher von vielen jahren geklagten grauaminum irung vn̄ Misuerstende zugetragen / Das zu entlicher entscheidung der selben / von des gnedigen Landesfürsten wegen / vff hochgedachtes Fürsten Herzogen Julij hieorige gnedige anordnung / Wir Augustus von der Assenburg / An stadt des Herrn alten Cantzlers Lt: Francisci Museltins / Weil derselbe Leibes Schwachheit halber baldt anfangs der hiesigen

Qualitätssicherung? DTAQ!

Braunschweig-Wolfenbüttel, Heinrich Julius von: Landtags Abschiedt [...] zwischen [...] Heinrichen Julio/ Postulirtem Bischoffen zu Halberstadt/ unnd Hertzogen zu Braunschweig und Lüneburgk/ [...] Und S.F.G. Landtschafft des Fürstenthumbs Braunschweig Wolffebüttelschen theils zu Saltzdam. Wolfenbüttel, 1604.



Informationen

Quelle: HAB
Publikationstyp: Monographie
Umfang: 138 Scans
ca. 214675 Zeichen
ca. 29155 Tokens
/[[:alnum:]]/
ca. 6636 Oberflächentypes

Schriftart: Fraktur
Genre:
im DTA seit: 2012-12-10 14:22:31
zuletzt geändert: 2013-11-11 11:47:04
Lizenz: Dieses Werk ist gemeinfrei.
Grundlage dieses Digitalisats: Obrigkeitskritik und Fürstenberatung: Die Oberhofprediger in Braunschweig-Wolfenbüttel 1568-1714: Bereitstellung der Texttranskription und Auszeichnung in XML/TEI. (2013-02-15T13:54:31Z) Bitte beachten Sie, dass die aktuelle Transkription (und Textauszeichnung) mittlerweile nicht mehr dem Stand zum Zeitpunkt der Übernahme entsprechen muss.

Wolfenbütteler Digitale Bibliothek: Bereitstellung der Bilddigitalisate (2013-02-15T13:54:31Z)
Marcus Baumgarten, Frederike Neuber, Frank Wiegand: Konvertierung nach XML gemäß DTA-Basisformat, Tagging der Titelblätter, Korrekturen der Transkription. (2013-02-15T13:54:31Z)

Weitere Informationen:

- Anmerkungen zur Transkription:
- Langes s (ſ) wird als rundes s (s) wiedergegeben.
 - Rundes r (ʀ) wird als normales r (r) wiedergegeben bzw. in der Kombination sc. als et (etc.) aufgelöst.
 - Die Majuskel J im Frakturdruck wird in der Transkription je nach Lautwert als I bzw. J wiedergegeben.
 - Übergeschriebenes „e“ über „a“, „o“ und „u“ wird

Ansichten

- Korrekturumgebung
- Faksimiles
- Download:
 - Text (UTF-8, Seitenumbrüche als ASCII \014)
 - TEI/XML (mit Silbentrennung)
 - HTML (mit Silbentrennung)
 - TCF (text annotation layer)
 - TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
 - TEI-Header
 - CMDI
 - Dublin Core
- Lemmata:
 - nach Frequenz
 - nach Frequenz (nur Nomen)
 - Wortwolke
 - Wortwolke (nur Nomen)
- Wortformen (Types):
 - nach Frequenz
 - nach Frequenz (nur Nomen)
 - Wortwolke
 - Wortwolke (nur Nomen)
- XML herunterladen und Werk sperren

Suche im Buch

?
 +

Formeln in diesem Buch

Korrekturstatus der Seiten

Text	Text/Bild
0 - 0 - 138	0 - 0 - 138
mind. zweimal kontrolliert, einmal kontrolliert, nicht kontrolliert	

offene Tickets – ganzes Buch

Zur Zeit gibt es keine offenen Tickets (ganzes Buch) zu diesem Buch.

offene Tickets

- #57091 [2013-08-27T09:16, FrankWiegand]
S. 3: Transkriptionsfehler: **freundlichem** freuntlichem
- #53124 [2013-06-26T11:16, KayMW]
S. 67: Transkriptionsfehler: **unterschiedlich**

Fazit

- Erfassung von Texten sollte nach dokumentierten Richtlinien und standardisierten Formaten erfolgen.
- Die Repositorien in der Cloud sollten ein erweitertes Angebot von Schnittstellen,
 - die erweiterte und genaue Metadaten,
 - die hochaufgelöste Bilder,
 - Transkriptionen von Editionen
 - sowie OCR-Daten anbieten.

Quellen

- Die auf den Folien erwähnten Firmen-, Software- und Hardwarebezeichnungen sind in den meisten Fällen auch eingetragene Warenzeichen und unterliegen als solche den gesetzlichen Bestimmungen. Alle erwähnten Markenrechte stehen dem jeweiligen Rechtsinhaber zu. GOOGLE sind eingetragene Warenzeichen der Google Corporation. ABBYY ist ein Warenzeichen der ABBYY Company. TextGrid ist ein vom Bundesministerium für Bildung und Forschung (BMBF) gefördertes Forschungsprojekt. Es besteht aus zehn Kooperationspartnern und wird koordiniert von der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB Göttingen). zeno.org ist eine Bezeichnung und Internetseite der Contumax GmbH & Co. KG. Wikisource ist eine Internetseite, Teil der Wikimedia foundation sowie ein Schwesterprojekt der Wikipedia.
- Folie 4: Friedrich II.: Über die deutsche Literatur [...]. Aus dem Französischen übersetzt. Berlin, 1780, S. 6. Online-Version: <http://www.deutschestextarchiv.de/friedrich/literatur/1780/viewer/image/text/40/7/>, Bild: <http://stabikat.de/DB=1/XMLPRS=N/PPN?PPN=137553501>
- Folie 5: <http://theartofgooglebooks.tumblr.com/>
- Folie 6: F. C. Theodor Krüger: Was ich eronnen, was ich erlebt! Gedichte. New York : Im Selbstverlage des Verfassers, 1897, S. 7, online-Version <http://hdl.handle.net/2027/wu.89068036045>
- Folie 13: http://de.wikisource.org/wiki/Unterhaltungen_deutscher_Ausgewanderten
- Folie 23: Das DTA-Zoning Tool ist ein Programm zur Markierung von Zonen auf Digitalisaten. <http://www.deutschestextarchiv.de/doku/software#ZOT>
- Folie 27: <http://diglib.hab.de/drucke/gn-6792-2s/start.htm>
- Folie 32: <http://www.deutschestextarchiv.de/dtaq/book/view/30178?p=3>
- Folie 33: <http://www.deutschestextarchiv.de/dtaq/book/show/30178>