

Deutsches Textarchiv

Deutsches Textarchiv. <http://www.deutschestextarchiv.de/> (Last Accessed: 2017-08-16). Reviewed by  Dario Kampkaspar (Herzog August Bibliothek / Austrian Centre for Digital Humanities an der Österreichischen Akademie der Wissenschaften), `dario [dot] kampkaspar [at] oeaw [dot] ac [dot] at`.

DOI: 10.18716/ride.a.6.8

Abstract

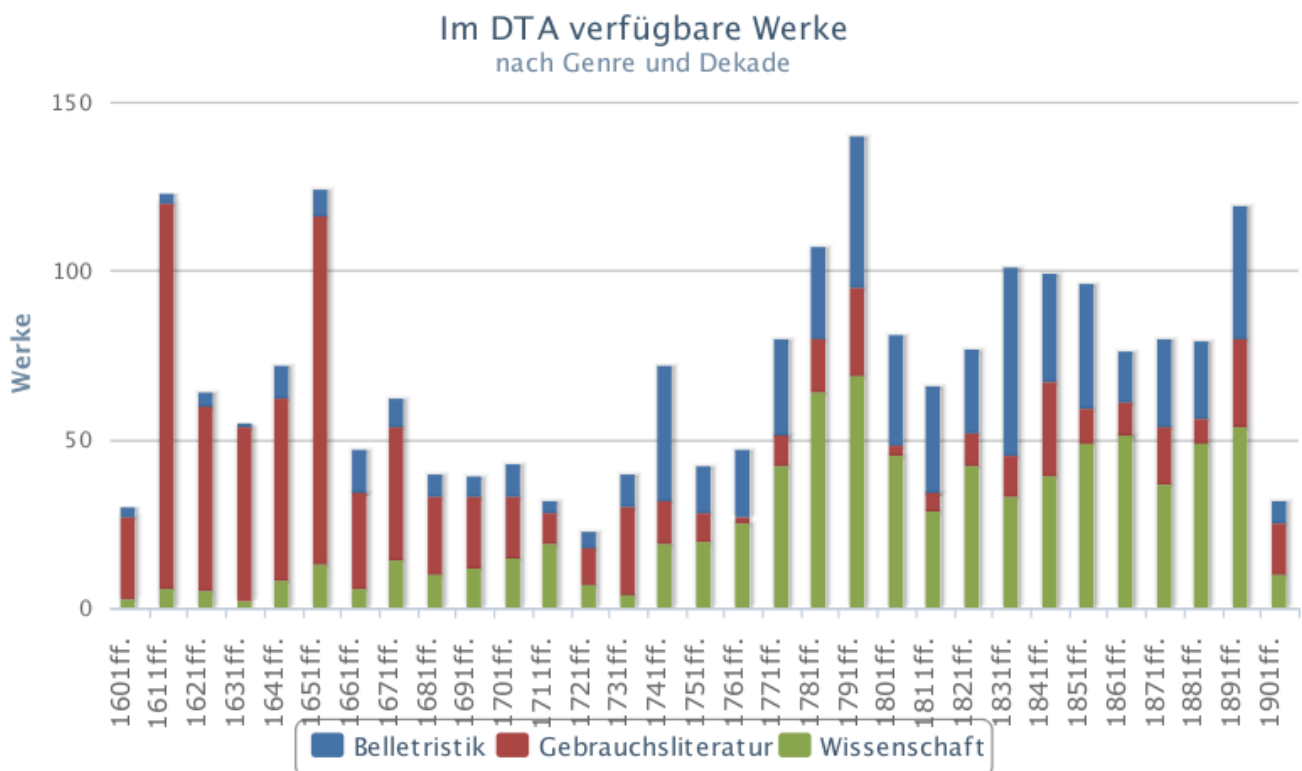
Owing to its well documented TEI subset and highly accurate transcriptions usually based on the first edition of a text, the German Textarchive (= Deutsches Textarchiv, DTA) is currently one of the best corpora for historical German texts (1600-1900), albeit not necessarily the most extensive. To date (August 2017) it contains around 3260 works. The full texts of the corpus are enhanced by digital facsimiles and encoded with regard to their visual features (e.g. layout, fonts) as well as annotated linguistically (e.g. PoS-tagging). The search engine focuses on linguistic features and allows for searching both exact spellings as well as spelling variants of a word. Due to its goal of providing a corpus for historical linguistics and the underlying selection criteria of the collection, no further comments or variants in other editions of one work than the first print are given. Even though there are minor points of criticism, due to the quality of its textual sources and the accurate documentation the DTA can be seen as a point of reference for other corpora.

Einleitung

Das Deutsche Textarchiv (DTA) entstand als Projekt an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) und wurde von 2007 bis 2016 von der Deutschen Forschungsgemeinschaft (DFG) unter der Leitung von Wolfgang Klein und Alexander Geyken gefördert. Mit derzeit (August 2017) über 620.000 Seiten aus mehr als 3.260 verschiedenen Werken aus der Zeit zwischen ca. 1600 und 1900* gehört das DTA zwar nicht zu den größten Textkorpora, doch dies ist gar nicht sein Anspruch. Die übergreifende Leitlinie wird dem Nutzer gleich auf der Startseite mitgeteilt: * „Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache“ oder, detaillierter, in den Leitlinien: * „Ziel des Deutschen Textarchivs (DTA) ist die Erstellung eines disziplinenübergreifenden Volltextkorpus deutschsprachiger Texte“. In diesem Sinne geht es mehr um die Erreichung und Einhaltung einer verlässlichen Qualität der Texterfassung, insbesondere was die sprachlichen Eigenarten eines Textes angeht. Neben dieses ‚Kernkorpus‘ treten noch die sogenannten DTA-Erweiterungen.* Hierbei handelt es sich um Texte, die teils unter anderen Gesichtspunkten ausgewählt und digitalisiert wurden (siehe Abschnitt zum DTA Erweiterungskorpus DTAE).

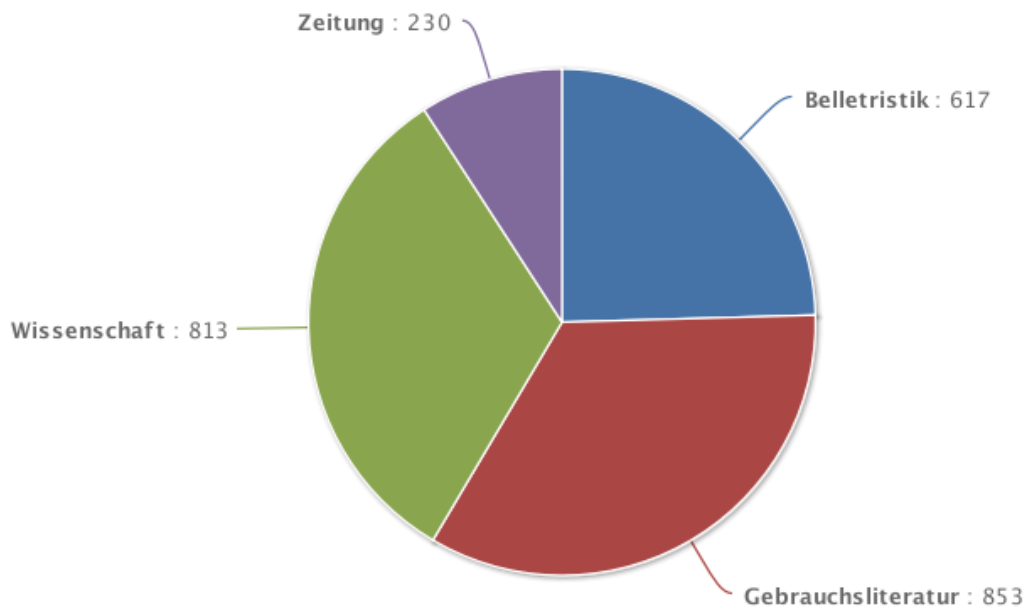
Das Kernkorpus

Aus der Zielsetzung des Kernkorpus ergeben sich viele der Eigenschaften des Deutschen Textarchivs, allen voran der Umstand, dass bei der Textauswahl und -aufbereitung hauptsächlich linguistische Gesichtspunkte im Vordergrund standen. Die Werke im DTA wurden nicht nach thematischen Gesichtspunkten ausgewählt, sondern, dem Ziel des Korpus entsprechend, wurde auf eine ausgewogene Zusammenstellung nach Textgenres geachtet und auf eine historisch-kritische Kommentierung der Texte verzichtet. Zur bestmöglichen Dokumentation des Sprachstandes wird überwiegend die erste verfügbare Ausgabe wiedergegeben.



Zeitliche Verteilung der Texte.

Im DTA verfügbare Werke nach Genre



Texte nach Genre.

Auch wenn der Anspruch einer ausgewogenen Textauswahl zur Abbildung des Neuhochdeutschen Sprachstandes und seiner Entwicklung nachvollziehbar ist, ist die Dokumentation, in der die Kriterien dieser Auswahl beschrieben werden,^{*} etwas dünn. So heißt es: Es wurden „mehrere Quellen herangezogen“, wobei „ausgewählte Literaturgeschichten“ und „Empfehlungen der Mitglieder der BBAW“ explizit genannt werden. Die Dokumentation der Textauswahl enthält zwei Graphiken: eine zur zeitlichen Verteilung der Werke (Abb. 1), die zur Zeit nur bedingt als gleichmäßig angesprochen werden kann, und eine zur Verteilung der Texte nach Genre (Abb. 2), wobei hier ‚Belletristik‘, ‚Gebrauchsliteratur‘, ‚Wissenschaft‘ und ‚Zeitung‘ genannt werden.^{*}

Scheint die Verteilung unter den ersten drei Gruppen noch recht gleichmäßig zu sein, stellt eine detailliertere Betrachtung^{*} ein etwas anderes Bild dar: So enthält die Gruppe ‚Belletristik‘^{*} beispielsweise eine Untergruppe ‚Briefe‘ mit mehreren Bänden Briefsammlungen^{*} (sowie einem Tagebuch!), während die Untergruppe ‚Novelle‘ deutlich mehr Werke verschiedener Personen umfasst (mit einem auffälligen Übergewicht auf den Werken Theodor Storms). Auch andere Untergruppen wie etwa ‚Lyrik‘; ‚Prosa‘ zeigen, dass hier gegebenenfalls noch etwas Klassifizierungs- wie Aufteilungsarbeit geleistet werden könnte.

Texterfassung und Formate

Das DTA versucht, bei den Texten des Kernkorpus eine möglichst hohe Erfassungsgenauigkeit zu erreichen. Neben der initialen Erfassung der Texte, die laut den Leitlinien in der Regel per *Double Keying* erfolgt, steht eine eigene browserbasierte Umgebung zur Qualitätskontrolle (DTA-Qualitätssicherung, DTAQ^{*}) zur Verfügung. Über DTAQ können registrierte Benutzer Korrekturen und Annotationen am Text anbringen, die dann von Mitarbeitern des DTA geprüft und gegebenenfalls in die Texte eingearbeitet werden. Detailliert ausgearbeitete und sehr gut dokumentierte Regeln zur Transkription erlauben eine sehr verlässliche

Auswertung der Texte, was auch durch das zugrunde gelegte Datenmodell unterstützt wird. Hierbei wurde basierend auf TEI P5 das DTA-Basisformat (DTA-Bf)* erstellt, das ebenfalls vorbildlich dokumentiert ist. Dieses Format wurde dementsprechend auch von der DFG als Format für historische Texte empfohlen.*

Der Fokus des Formates und der Erfassungsarbeit liegt auf der exakten Erfassung der Textstruktur sowie der originalen Schreibweisen und damit dem Aufbau eines Korpus historischer Wortformen. Für die weitergehende linguistische Bearbeitung wird auf das im Kontext von CLARIN-D entwickelte Text Corpus Format (TCF)* zurückgegriffen.

Wie bereits angesprochen kann die Dokumentation der Texterfassung und Modellierung des DTA in weiten Teilen als vorbildlich angesehen werden. Neben verschiedenen thematischen Zugängen (so zum Beispiel Metadaten und formale wie inhaltliche Erschließung) gibt es Übersichten zu den verwendeten Elementen und eine Suche innerhalb der Dokumentation. Da die Einhaltung dieser Regeln im Rahmen der Qualitätskontrolle DTAQ gesichert wird, sind die Texte des Kernkorpus in vielen Belangen der Goldstandard im Hinblick auf die weitere Verwendbarkeit. Ebenfalls positiv anzumerken ist, dass auch die Bestandteile der Dokumentation durch die Lizenz CC BY-SA 3.0 frei verfügbar sind.

Generelle Benutzerführung

Der Zugriff auf die Texte ist auf mehreren Wegen möglich. Am auffälligsten ist dabei direkt auf der Startseite die Suchmöglichkeit, die bei einer Volltextsuche innerhalb des Korpus auch linguistische Verfeinerungen erlaubt.

Als linguistische Suchengine wird DDC („Dialing/DWDS-Concordancer“) verwendet, das umfangreiche linguistische Abfragen bietet. Unterstützt wird diese durch CAB („Cascaded Analysis Broker“), ein Tool zur Findung moderner Wortformen zu historischen Schreibungen. Auch die (im Vergleich zu einfachen Volltextsuchen natürlich umfangreichere) Syntax der Engine ist mit Beispielen dokumentiert.* Einige Features sind dabei die lexembasierte Suche (durch CAB gefundene Formen zu einem Grundlexem), Suche nach einer genauen Form mit oder ohne Trunkierung und die Suche von Phrasen (genaue Phrasen oder auch Wörter mit bestimmten Abständen zueinander) oder eine Suche aufgrund der Wortart. Diese Suchmöglichkeiten können für komplexe Anfragen miteinander kombiniert werden.

Anmelden (DTAQ) DWDS [diesDB](#) [CLARIN-D](#)

DTA "wegen @dem" [suchen](#) [Hilfe](#)

in den Titeldaten ☒ im Korpus ☐ in der Dokumentation

Suche im Deutschen Textarchiv

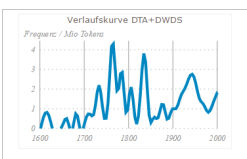
Treffer 1 - 10 von 193

Neue Suche · KWIC [suchen](#) [Hilfe](#)

10 Treffer pro Seite Sortierung: Datum aufsteigend/absteigend · zufällig

gehe zu: Anfang · -10 · -5 · vorherige · nächste · +5 · +10 · Ende

- 1: Arnim, Bettina von: Goethe's Briefwechsel mit einem Kinde, Bd. 1, Berlin, 1835, #362
Wir gingen mit einem irreführenden Wegweiser durch eine Thalschlucht einem Fluß entlang, den man die Wisper nennt, wahrscheinlich **wegen dem** Rauschen des Wassers, das über lauter platte Felssteine sich windet, und in den Lücken schäumt und flüstert. [mehr]
- 2: Baer, Karl Ernst von: Über Entwicklungsgeschichte der Thiere, Bd. 1, Königsberg, 1828, #175
Da die Wirbelsaite der dünnste Theil ist, den man findet, so hat es keine Wahrscheinlichkeit, dass im Embryo Theile vorkommen, die ihrer Dünne **wegen dem** Microscope gar nicht erreichbar wären. [mehr]
- 3: Arnim, Bettina von: Die Günderode, Bd. 2, Grünberg u. a., 1840, #268
Aber der Ephraim liegt mir am Herzen daß der jetzt wo die Natur schläft und nur aufrührische Träume hat, die eisige bergige Straße wendert, wo es so früh Nacht ist und wo er in schlechte Herbergen kommt; aber er sagt er habe einen Tag schon versäumt **wegen dem** Wetter, und seine Enkel warten alle auf ihn, die würden so schon in großen Sorgen um ihn sein, und das Sturmwetter werde er schon ertragen, er habe es schon mehr mitgemacht, und sein Enkel trägt den Bündel. [mehr]
- 4: Brümmer, Franz: Lexikon der deutschen Dichter und Prosaisten vom Beginn des 19. Jahrhunderts bis zur Gegenwart. Bd. 2, 6. Aufl. Leipzig, 1913, #218
Hier lebt sie noch jetzt, hat aber ihrer Krankheit **wegen dem** Schreiben fast ganz entsagen müssen. [mehr]
- 5: Cancrin, Franz Ludwig von: Beschreibung der vorzüglichsten Bergwerke, Frankfurt (Main), 1767, #65
Er ist, **wegen dem** Trieb sand, ausgemauert, und mit ovalen Lichtlöchern versehen worden, die alle ausgemauert sind, damit man denselben um desto eher auf erhalten, und nicht nötig haben möge, beständig an ihm zu verzimmern. [mehr]
- 6: Cancrin, Franz Ludwig von: Beschreibung der vorzüglichsten Bergwerke, Frankfurt (Main), 1767, #95
Denn wann die Schiefer erst einmal in dem Brand sind: So brennen sie **wegen dem** in ihnen befindlichen Schwefel, in sich. [mehr]



Fahren Sie über die einzelnen Tokens mit der Maus, um folgende Informationen zu sehen:

- u: Originaltext, UTF-8-kodiert
- w: approximierter Latin-1-Text
- v: CAB-normalisierte Wortform
- l: Lemma (unklutierte Form)
- pi: Part-of-Speech-Analyse

Ergebnisansicht der Suche.

In der Ergebnisansicht (Abb. 3) werden die Treffer im Kontext ihres Satzes angezeigt, wobei jeder Eintrag in der Liste einem Treffer entspricht (mehrere Treffer auf der gleichen Seite erzeugen standardmäßig jeweils eigene Einträge).^{*} Ebenfalls angeboten wird eine Verlaufskurve der gesuchten Form innerhalb des (gewählten Teil-)Korpus, wobei die Zahl der Treffer je einer Million Tokens über der Zeit angegeben wird. Ein Klick auf diese Graphik führt zu einer detaillierteren Ansicht.

Neben der Volltextsuche gibt es in der Seitenspalte der Startseite die Möglichkeit zum ‚Stöbern im DTA‘, unter der alle Texte nach dem Jahrhundert der Veröffentlichung sowie nach der Textgattung (hier sehr weit gefasst in die drei Gruppen ‚Belletristik‘, ‚Gebrauchsliteratur‘ und ‚Wissenschaft‘; eine feinere Aufteilung erfolgt auf den jeweiligen Unterseiten) gruppiert werden. Dazu kommt im Menü ‚Texte‘ eine Zeitleiste, die die Titel nach ihrem Erscheinungsjahr zusammengefasst aufführt. Wer Werke eines bestimmten Verfassers sucht, muss entweder die Suche in den Titeldaten bemühen oder – weniger offensichtlich – in der Seitenspalte unter ‚Neue Werke im DTA‘ dem Link ‚alle Titel ...‘ folgen.

Textansichten

Anmelden (DTAQ)

DTA

suchen

Hilfe

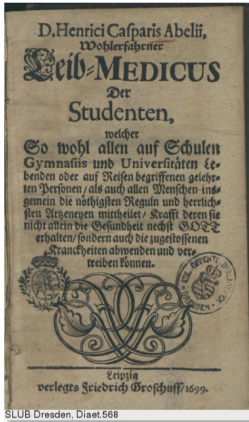
in den Titeldaten

im Korpus

in der Dokumentation

DWDSlexDBCLARIN-D

TexteProjektDokumentationImpressum



SLUB Dresden, Diæt.568

Text-Bild-Ansicht öffnen ...

Abel, Heinrich Kaspar: Wohlerfahner Leib-Medicus der Studenten. Leipzig, 1699.

BIBLIOGRAPHISCHE ANGABEN

URN:urn:nbn:de:kobv:b4-200905199878

Titel:Wohlerfahner Leib-Medicus Der Studenten

Untertitel:welcher So wohl allen auf Schulen Gymnasii und Universitäten Lebenden oder auf Reisen begriffenen gelehrten Personen/ als auch allen Menschen insgemein die nöthigsten Reguln und herrlichsten Artzeneyen mittheilet/ Krafft deren sie nicht allein die Gesundheit nechst Gott erhalten/ sondern auch die zugestossenen Kranckheiten abwenden und vertreiben können

Autor/in:Heinrich Caspar Abel (GND, Wikipedia, ADB/NDB)

Erscheinungsjahr:1699

Verlag:Groschuff

Ort:Leipzig

Auflage:1. Auflage

Bibliothek:SLUB Dresden

Signatur:SLUB Dresden, Diæt.568

INFORMATIONEN ZUM WERK

Publikationstyp:Monographie

Verfügbarkeit:Text (TEI-XML-, HTML-, E-Book-Fassung): CC BY-NC 3.0
Weitere Informationen: Nutzungsbedingungen.

Schriftart:Fraktur

Genre:Wissenschaft :: Medizin

Im DTA seit:2011-01-28 20:05:53

Korpus:DTA-Kernkorpus

GRUNDLAGE DIESES DIGITALISATS

Dieses Werk wurde gemäß den DTA-Transkriptionsrichtlinien im Double-Keying-Verfahren von Nicht-Muttersprachlern erfasst und in XML/TEI P5 nach DTA-Basisformat kodiert.

INHALTSVERZEICHNIS

[Titelseite]

Vorrede.

Register derer Capitel.

Studenten-Kranckheiten und Derer Curen Ersten Buchs erste Abtheilung

Ersten Buchs/ anderer Theilung/

Andern Buchs erster Theilung.

Andern Buchs anderer Theilung.

Dritten Buchs erster Theilung. Von der Studenten Diæt.

Dritten Buchs anderer Theilung Von der Luft insgemein.

Dritten Buchs dritter Theilung.

Dritten Buchs 4. Theilung. Von andern Stück der Diæt, nemlich von dem Tranck.

Dritten Buchs 5. Theilung.

Dritten Buchs 6. Theilung. Von Gemüths-Bewegung.

Dritten Buchs 7. Theilung. Vom Schlaffen und Wachen.

Dritten Buchs achter Theilung.

Dritten Buchs neuer Theilung Wie man sich verhalten sol in den vier Jahres Zeiten.

Viertes Buch/ Erster Theilung. Wie die Studenten Kranckheiten zu curiren.

Hinweis: Dieses Inhaltsverzeichnis wurde automatisch aus den XML-Quellen erstellt.

URL zu diesem Werk: http://www.deutschestextarchiv.de/abel_leibmedicus_1699

Zitationshilfe: Abel, Heinrich Kaspar: Wohlerfahner Leib-Medicus der Studenten. Leipzig, 1699. In: Deutsches Textarchiv <http://www.deutschestextarchiv.de/abel_leibmedicus_1699>, abgerufen am 26.05.2017.

Suche im Werk

suchen

Hilfe

Ansichten für dieses Werk

Text-Bild-Ansicht

alle Faksimiles

DTAQ (Qualitätssicherung)

Download

XML (TEI P5) · HTML · Text

TCF (text annotation layer)

TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)

Metadaten

TEI-Header · CMDI · Dublin Core

Statistiken

Scans:

412

Zeichen:

ca. 350 279

Tokens:

ca. 49 972

Oberflächentypen:

ca. 11 323

Wortwolken

Lemmata

Lemmata (nur Nomen)

Types

Types (nur Nomen)

Voyant Tools

transliterierter Text

normalisierter Text

lemmatisierter Text

Einstiegsseite eines Textes.

Die Einstiegsseite eines Textes (Abb. 4) bietet bibliographische Angaben, weitere Metadaten, den Zugang zu den Ansichten des Textes, verschiedene Download-Optionen sowie Statistiken und Links zu einigen Tools (zur Zeit Wortwolken sowie die Übergabe des transliterierten, normalisierten oder lemmatisierten Textes an *Voyant Tools*). Die bibliographischen Angaben enthalten die üblichen Titelangaben, die URN des Textes zur permanenten Adressierung sowie Informationen zu dem für die Transkription verwendeten Exemplar. Die Metadaten sind neben anderen Formaten als TEI-Header und Dublin Core verfügbar, während für den Volltext verschiedene Versionen angeboten werden: erfasster Volltext als XML (im DTA-Bf), eine HTML-

DOI: 10.18716/ride.a.6.8

Page 6

Ansicht, Plain Text, sowie linguistische Annotationen verschiedenen Umfanges als TCF. Die Download-Option ‚HTML‘ gibt zwar den gesamten Text als HTML wieder, bietet diesen aber tatsächlich zum Speichern an; den Volltext in einer einzigen HTML-Datei online einzusehen, scheint nicht vorgesehen zu sein.

The screenshot shows the DTA (Deutsches Textarchiv) interface. At the top, there's a navigation bar with 'Anmelden (DTAQ)', search bars, and links to 'DWDS', 'denDB', and 'CLARIN-D'. Below this, the document title 'Abel, Heinrich Kaspar: Wohlerfahner Leib-Medicus der Studenten. Leipzig, 1699.' is displayed. The main content area is split into two panes. The left pane shows a scanned image of the title page of the book, with the title 'D. Henrici Casparis Abellii, Wohlerfahner Leib-Medicus der Studenten, welcher So wohl allen auf Schulen Gymnasiis und Universitäten Lebenden oder auf Reisen begriffenen gelehrten Personen/ als auch allen Menschen-insgemein die nöthigsten Regeln und herrlichen Artzneyen mittheilet/ Krafft deren sie nicht allein die Gefundtheit nechst GOTT erhalten/ sondern auch die zugestoffenen Kranckheiten abwenden und vertreiben können.' The right pane shows the transcribed text in a structured HTML format, including the title, author, and a detailed preface. The interface includes search bars, navigation controls, and a sidebar with metadata and download options.

Parallelansicht einer Seite.

Der Standard-Zugang zum Text ist die Text-Bild-Ansicht, die eine seitenweise Parallelansicht bietet (Abb. 5). Die auf der Einstiegsseite gebotenen Optionen sind hier (bei ausreichender Fensterbreite) am rechten Seitenrand zu finden. Die parallele Ansicht stellt Digitalisat und Transkription nebeneinander, wobei für die Transkription verschiedene Anzeigeformate gewählt werden können: das originale XML, eine HTML-Ansicht der getreuen Abschrift oder verschiedene Normalisierungen. Leider sind die Größen der Seitenbestandteile fix vorgegeben, sodass teils sogar bei kurzen Seiten innerhalb der Transkription gescrollt oder die Schriftgröße angepasst werden muss, um den gesamten Text einer einzelnen Textseite zu lesen. Die Informationen zum gesamten Werk sind auch hier jederzeit abrufbar und es werden außerdem der Link zum gesamten Werk, zur einzelnen Seite (nicht aber jeweiligen Ansicht) und eine Zitationshilfe mitgegeben. Verschiedene Versionen des Textes, der sich durch die Qualitätssicherung DTAQ ändern könnte, sind nicht abrufbar. Zwar ist durch die bereits vor der Veröffentlichung greifenden Kontrollmechanismen nicht von großen Änderungen auszugehen, doch ist gerade bei den oft feinfühligsten linguistischen Suchen hier das Problem der fehlenden Nachvollziehbarkeit einer Aussage, die auf einem älteren Datenstand basiert, nicht

von der Hand zu weisen. Bislang noch nicht umgesetzt, aber eine denkbare Erweiterung wären, sind Möglichkeiten zur Anbringung privater Annotationen oder Lesezeichen sowie Möglichkeiten zur Adressierung von Textauszügen, zum Beispiel unter Nutzung der Tokenisierung.

DTA Erweiterungskorpus (DTAE)

Digitalisierungsprojekte an, insbesondere bei den folgenden Aufgaben:

- Bilddigitalisierung**
Wir können Ihnen den Kontakt zu den Bibliotheken herstellen. Das DTA kooperiert mit verschiedenen Bibliotheken, deren Sammelschwerpunkte bei deutschen Drucken des 17. bis frühen 20. Jahrhunderts liegen.
- Erfassung der Metadaten**
Sowohl bibliographische als auch spezifische Daten zum digitalen Faksimile sowie zur Textfassung sind notwendig. Für die grundlegenden Metadaten zum Text stellt das DTA eine spezielle Formular zur komfortablen Erstellung eines basisformatkonformen TEI-Headers zur Verfügung.
- Volltextdigitalisierung**
Das DTA kann die Digitalisierung Ihrer Texte entsprechend den DTA-Richtlinien übernehmen oder diesen Arbeitsschritt beratend unterstützen. Für die Textfassung bietet DTAE erprobte und ausführlich beschriebene Erfassungsrichtlinien. Die Dokumentation des DTA-Basisformats veranschaulicht mit Beispielen, wie das XML-Markup verwendet wird. DTAE bietet Ihnen darüber hinaus spezifische Anpassungen für den XML-Editor oXygen und das Textverarbeitungsprogramm Microsoft Word. Hiermit soll Ihnen die Transkriptions- und Annotationsarbeit unter Anwendung der DTA-Richtlinien erleichtert werden.

TEXTQUELLEN IN DTAE

nach Seiten gruppiert

Insgesamt wurden durch DTAE bisher 164490 Seiten in das DTA aufgenommen.

4 KOOPERATIONEN IM RAHMEN VON DTAE

- Alexander-von-Humboldt Forschungsstelle (BBAW): Digitalisierung unselbstständig erschienener Schriften Alexander von Humboldts
- Forschungsstelle für Personalschriften (AdW Mainz): Digitalisierung von Leichenpredigten aus dem Bestand der ehemaligen Stadtbibliothek Breslau
- Herzog August Bibliothek Wolfenbüttel, Wolfenbütteler Digitale Bibliothek (WDB), Projekt AEDIt Frühe Neuzeit: Integration von Volltexten aus verschiedenen Projekten der HAB/WDB
- Johann Friedrich Blumenbach – online (AdW zu Göttingen): Integration von deutschsprachigen Texten Blumenbachs in das DTA und Bereitstellung einer Rechercheschnittstelle
- Marx-Engels-Gesamtausgabe (BBAW): Aufbereitung und Publikation der digitalisierten

WIKISOURCE^[1]

Werke anzeigen ...

Novellenschatz^[1]

Werke anzeigen ...

NTSM^[1]

Werke anzeigen ...

SBB Funeralschriften^[1]

Werke anzeigen ...

TdeF^[1]

Werke anzeigen ...

Wikisource^[1]

Werke anzeigen ...

Ausschnitt der Seite zum DTAE.

Die derzeit (August 2017) knapp 165.000 Seiten im Erweiterungskorpus* stammen aus externen Quellen, wobei es sich um Korpora einzelner Institutionen wie der Wolfenbütteler Digitalen Bibliothek (WDB), der Herzog August Bibliothek Wolfenbüttel (HAB), Kooperationsprojekte wie ‚AEDIt‘ (HAB, BBAW und Akademie der Wissenschaften und Literatur in Mainz) oder Einzelprojekte wie ‚Semantic Blumenbach‘ an der Akademie der Wissenschaften in Göttingen handeln kann (Abb. 6). Ergänzt wird das DTAE durch weitere Volltexte, die aus digitalen Archiven (überwiegend Wikisource) übernommen und in das DTA-Bf überführt wurden.

Ausführlichere Informationen zu den einzelnen Quellen der übernommenen Texte sind nicht auf derselben Seite zu finden – unter der Graphik sind nur kurze Informationen zu den Kooperationen aufgeführt –, sondern unter ‚Textquellen‘ in der Dokumentation.* Hier erst findet man die Erklärung, dass sich unter dem Beiträger ‚DTAE‘ (wie er in der Grafik in Abb. 6 auftritt) kein Zirkelschluss verbirgt, sondern eine Zusammenfassung kleinerer Einzelquellen ohne genaue Zuordnung zu einer Institution.

In der rechten Spalte lässt sich auch eine Übersicht der Texte aufrufen, die aus den einzelnen Projekten übernommen sind. Eine Angabe, zu welchem Zeitpunkt und in welcher Form (Bilddigitalisat, Volltexte) Daten von einer Quelle übernommen wurde, sucht man hier wie auch in der Kurzbeschreibung vergebens. Die

Liste enthält alle übernommenen Texte, auch wenn sie noch nicht kontrolliert wurden. Es ist nicht erkennbar, welche Links zu einem Volltext führen und welche zum DTAQ. Eine deutlichere Kennzeichnung wäre für den Benutzer hilfreich.

Zugänglichkeit und Nachnutzung

Die Texte werden in der Regel unter verschiedenen freien Lizenzen zur Verfügung gestellt. Genauere Angaben lassen sich auf den jeweiligen Seiten einzelner Werke entnehmen. Die verschiedenen zur Verfügung gestellten Downloadformate und Lizenzen erlauben eine Weiterverwendung in verschiedenen Forschungskontexten.

Als einfachste Variante der Übernahme ist ein Download des gesamten Korpus (nur Kernkorpus oder Kern- und Ergänzungskorpus) wie auch der einzelnen Gruppen, wie sie unter ‚Stöbern‘ zu finden sind, möglich. Eine Gruppierung der Texte durch den Nutzer ist zurzeit nicht umgesetzt. Es werden außerdem einzelne ‚Versionen‘ dieser Download-Optionen angeboten, wenngleich diese eher den Charakter von Snapshots mit schwankendem zeitlichen Abstand von teils fast zwei Jahren haben und so nur als grobe Versionierung der Texte dienen können.

Die bei der Erstellung des Korpus verwendete Software wird im Rahmen der Dokumentation ebenfalls knapp beschrieben,^{*} wobei aber nicht immer klar ist, ob und inwiefern es sich um Eigenentwicklungen handelt. Teile der Software sind für die eigene Nutzung verfügbar, doch ist bei den externen Links oft nicht ersichtlich, in welchem Umfang und unter welcher Lizenz diese Software-Stücke genutzt werden können. Für die Verwendung im XML-Editor oXygen wird eine auf das DTA-Bf zugeschnittene Anpassung zur Verfügung gestellt, die mit einer Veröffentlichung im November 2013 aber auf einer nicht mehr aktuellen Version von oXygen basiert.

Erfreulich ist das Angebot an APIs. Neben einer OAI-PMH Schnittstelle wird auch eine API zu CAB angeboten.^{*} Etwas dürftig ist allerdings die ‚Dokumentation‘ der APIs ausgefallen, beschränkt sie sich doch auf eine einfache Aufzählung von Links. Ohne weitere Unterscheidung stehen hier Atom-Feeds neben einem Link zur OAI-PMH-Schnittstelle oder einem OpenSearch-Wrapper für die oben beschriebene Suche. Ergänzende Informationen könnten an dieser Stelle hilfreich sein.

Abschließende Bemerkungen

Das Deutsche Textarchiv bietet seinem Anspruch entsprechend qualitativ hochwertig aufbereitete Texte, die insbesondere, aber nicht nur für die linguistische Nachnutzung einen ausgezeichneten Standard bilden. Die Dokumentation vor allem des verwendeten Datenformates kann mit Fug und Recht als vorbildlich bezeichnet werden. Kleinere Eigenarten der Benutzerführung oder der zur Verfügung gestellten Ansichts-, Browsing- oder Downloadoptionen stellen keine gravierenden Probleme dar und werden vom Projektteam im Rahmen ihrer Möglichkeiten sicherlich auch angegangen werden.

Es steht zu hoffen, dass auch in Zukunft das hohe Niveau der Textaufbereitung gehalten und das Textkorpus sukzessive weiter ausgebaut werden kann. Leider ist der Webseite nicht zu entnehmen, wie sich die weitere Zukunft des DTA gestalten wird, nachdem die Förderung durch die DFG 2016 ausgelaufen ist. Genauso fehlen Angaben, ob die Vorhaltung der Texte dauerhaft gesichert ist. Zwar ist davon auszugehen, dass ein solch prominentes Projekt einer Akademie nicht sang- und klanglos verschwinden wird, doch sollten nach Ablauf eines Projektes zumindest kurze Hinweise zu Fragen der Archivierung und langfristigen Verfügbarkeit gegeben und möglichst sichtbar platziert werden.

Der weitere Ausbau des Korpus auch durch externe Kooperationspartner lässt hoffen, dass sich die Datengrundlage noch deutlich erweitern wird. Kleinere hier angeführte Vermutstropfen schmälern die Leistung nicht und werden durch das Projektteam sicher geprüft werden. Durch die Qualität der Texte, aber insbesondere auch durch die Dokumentation des DTA-Basisformates hat das DTA Vorbildcharakter für viele andere Projekte.

References

DTA: Deutsches Textarchiv. 2007-2017. Berlin-Brandenburgische Akademie der Wissenschaften. <http://deustchestextarchiv.de/>.

Figures

Zeitliche Verteilung der Texte.

Texte nach Genre.

Ergebnisansicht der Suche.

Einstiegsseite eines Textes.

Parallelansicht einer Seite.

Ausschnitt der Seite zum DTAE.

Notes

Die Angaben sind der auf der Startseite recht prominent platzierten Statistik entnommen, die einen guten Überblick über die laufende Arbeit am Korpus gibt.

DTA, Startseite: <https://web.archive.org/web/20170830180801/http://deustchestextarchiv.de/>.

DTA, Leitlinien: <https://web.archive.org/web/20170526105849/http://www.deustchestextarchiv.de/doku/leitlinien>.

DTA, DTAE: <http://www.deustchestextarchiv.de/dtae>.

DTA, Dokumentation: <https://web.archive.org/web/20170526085322/http://www.deustchestextarchiv.de/doku/textauswahl>.

„Zeitung“ als „Genre“ taucht allerdings in der gattungsorientierten Auswahl auf der Startseite nicht auf.

Eine Übersicht über die Texte der Gruppen ist allerdings von hier aus nicht möglich; sie kann stattdessen von der Startseite aus erreicht werden.

DTA, Genre Belletristik <https://web.archive.org/web/20170526103123/http://www.deustchestextarchiv.de/list/browse?genre=Belletristik>.

Unter anderem Bettina von Arnims „Goethes Briefwechsel“, eher ein Briefroman als eine unverfälschte Briefsammlung, und Ludwig Börnes „Briefe aus Paris“, die auf Basis der früheren Edition der BBAW wiedergegeben wurden.

DTA, DTAQ: <http://www.deustchestextarchiv.de/dtaq>.

DTA, Basisformat: <https://web.archive.org/web/20170526111458/http://www.deustchestextarchiv.de/doku/basisformat/>.

http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf und [http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/foerderkriterien_editionen_literaturwissenschaft.pdf](http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/foerderkriterien_editionen_literaturwissenschaft.pdf).

Weblicht, TCF-Format: http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format.

DTA, DDC-Suche: https://web.archive.org/web/20170526123024/http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe.

Nach der Dokumentation der Suche sollten die Ergebnisse „alphabetisch nach dem Titel der Werke, in dem die Ergebnisse gefunden wurden, sortiert“ werden (vgl. https://web.archive.org/web/20170821082531/http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe). Die Kopfzeile einer einfachen Suche nennt eine Sortierung nach Datum. Ohne weitere Eingriffe scheint jedoch, wie im abgebildeten Beispiel zu sehen, keine dieser Aussagen korrekt zu sein. Über die auf der Seite gegebenen Links oder über Parameter in der Suche kann zumindest eine Sortierung nach Datum erzwungen werden, eine alphabetische Anordnung ließ sich im Test nicht erreichen.

DTA, DTAE: <http://www.deutschestextarchiv.de/dtae>.

DTA, Textquellen: <https://web.archive.org/web/20170526092945/http://www.deutschestextarchiv.de/doku/textquellen>.

DTA, Software: <https://web.archive.org/web/20170526122208/http://www.deutschestextarchiv.de/doku/software>.

DTA, Dokumentation der APIs: <https://web.archive.org/web/20170529084103/http://www.deutschestextarchiv.de/api>, API zu CAB im Detail: <http://odo.dwds.de/~moocow/software/DTA-CAB/>.

Licence: <http://creativecommons.org/licenses/by/4.0/>

Published: 2017-09 · build cb141fb (2026-08-23T11:45:27+02:00)